A CLASSIFICATION OF LARGE AMERICAN URBAN AREAS

PREPARED FOR THE NATIONAL SCIENCE FOUNDATION

EMMETT KEELER
WILLIAM ROGERS

R-1246-NSF
MAY 1973

25th Year

Rand
SANTA MONICA, CA. 90406

This report was sponsored by the National Science Foundation under Grant GI-29763. Reports of The Rand Corporation do not necessarily reflect the opinions or policies of the sponsors of Rand research.

Published by The Rand Corporation

A CLASSIFICATION OF LARGE AMERICAN URBAN AREAS

PREPARED FOR THE NATIONAL SCIENCE FOUNDATION

EMMETT KEELER
WILLIAM ROGERS

R-1246-NSF
MAY 1973

Rand
SANTA MONICA, CA. 90406

PREFACE

The work reported here was done for the Rand Urban Policy Analysis Program, which is sponsored by the National Science Foundation. The Urban Program has concentrated its efforts on research and analysis in three American metropolitan areas: San Jose, Seattle, and St. Louis. The purpose of this report is to aid analysts in generalizing their results from these cities to others, and in selecting suitable cities for future study. Although the work tried specifically to determine policy relevant differences and similarities among American cities, the data collected for this purpose have proved useful in other cross-sectional studies, including the continuing analysis of central city decline.

Emmett Keeler is a member of the Rand Economics Department, and William Rogers is a consultant to The Rand Corporation.

SUMMARY

This study presents a new classification of the 125 largest American metropolitan areas. Its purpose is to help the Rand Urban Studies Group make generalizations about the country as a whole from the detailed studies of specific cities. Based on a data file of important structural variables, the classification shows how representative an urban area is and is thus useful in weighing evidence and in selecting further areas for study. The data file is an ongoing resource that is being used in cross-sectional testing of theories—for example, in Appendix D a statistical model is presented of the relationship between population and economic growth.

The specific variables used here were chosen from the experience of earlier Rand urban studies, which focused on growth and decline of areas, social and racial inequality, government and political structure, and city-suburban relationships. The resulting 53 variables measure these themes and the urban geography that is their setting. Tables of the range of these variables show the enormous diversity of American urban areas.

Factor analysis produced eight synthetic variables or "factors," which contain much of the variation of the 53 original variables. These factors are: manufacturing, income inequality, ghetto-suburb contrasts, growth, age of population, unemployment, density, and education. Instead of 53 variables, each city may be represented by its eight factor scores.

The metropolitan areas were divided into clusters of cities "similar" in these eight dimensions. Two types of results are presented: a complete list of the cities divided into ten clusters, where each cluster can be characterized by its "most representative" city—Columbia, San Diego, South Bend, Knoxville, Dallas, Worcester, Cleveland, Oxnard, San Antonio, and Pittsburgh; and a tree of the cluster types where the cities are divided into 4, 6, 8, 10, and 14 clusters. The tree shows how the larger clusters subdivide. For example, the Southern cities cluster is divided into mature cities (Columbia) and newer cities without very distant suburbs (Knoxville); these newer cities divide in turn into those with stable (Charlotte) and declining (Chattanooga) populations. *Although there are no regional variables per se, the clusters generally represent geographic regions, since regions tend to share a common history and economic development, which is reflected in the data.*

This work is supplemented by recent literature described in Appendix C. It has been shown, for example, that the definition of urban area—whether legal city, urbanized area, or standard metropolitan statistical area—does not affect the results of most statistical analyses. Also, as pointed out by other studies, factor analysis has

great value in exploratory work: Although there is ambiguity about what the factors mean, they have mathematical properties that save time and reduce exploratory problems. Appendix B discusses at greater length the philosophy and method of exploratory as opposed to classical data analysis. Accurate probability statements are waived for the chance to let expert judgment and new findings interact with the data.

ACKNOWLEDGMENTS

The authors would like to thank Vreni Keeler and Rand colleagues Julie DaVanzo and Dan Relles, for helpful comments and suggestions.

CONTENTS

I. INTRODUCTION

A primary objective of the Rand Urban Policy Analysis Program is to generalize about national urban problems and policies. However, operational considerations require most of the effort to be in-depth studies of a small number of cities.[1] The problem, then, is how to use the knowledge gained from these studies to make general conclusions on policy for the large and diverse set of American cities. The work reported here is designed to help in that generalization.

We will briefly describe our basic strategy for generalization. Hypotheses are formed in the course of in-depth studies of individual cities. These hypotheses are then tested on other intensively studied cities. If the results are consistent, they can be extrapolated by means of statistical procedures on data from the entire set of large American cities.

There are many differences between this strategy and that of the traditional scientific experiment. These differences are unavoidable, given our aims and the subject under investigation, but they should be acknowledged. First, because of the extreme complexity and diversity of city settings and interactions and the difficulty of measuring them, no strict scientific control procedures are possible. We intend to draw inferences that are plausible and consistent with our observations, but we do not expect that our findings can be "proved." Another difference lies in the transient nature of urban knowledge—the phenomena under investigation are changing rapidly, so that many of the better policies of today may be useless 20 years from now. A third difference is caused by our current lack of knowledge as to what is important. In contrast to scientific verification, in which the theory is set out in advance and the good experiment has no surprises, our work is sequential. As new hypotheses are formed or new important variables discovered, we try to improve our explanations. New analysis may suggest further new theories. We widen the scope of our inquiry at some expense of certainty in the results.

This work is designed to help generalization in three ways. First, we have created a data file of variables that seem important in theory and in the specific cities studied. The file can be used directly in cross-sectional testing of theories developed for the individual cities studied; for example, in Appendix D we give such an analysis of the relationship between economic and population growth in major urban areas. Second, the city classification reported here is useful in selecting the

[1] To date, the cities of San Jose, Seattle, and St. Louis have been studied. The work on San Jose is complete, but work continues on Seattle and St. Louis.

cities to be studed intensively. Since we can choose only a limited number of cities to study in detail, our aim is to pick cities that will be representative of most types of American cities. Although a balanced experimental design is impossible because of the large number of relevant parameters and our incomplete knowledge of which are most important, if our findings hold up on a wide variety of cities, then they are plausibly universal. The final aid to generalization lies in the weighing of evidence: How idiosyncratic is Seattle or St. Louis? If certain phenomena are present there, where else might they appear? By highlighting what appear to be the important parameters of city variation, we can get a better idea of the range and limitations of policy conclusions based on evidence from a few cities.

Since Thorndike's pioneering work in the 1930s, there have been many attempts at city classification.[2] In Appendix C, we discuss how that literature complements our own work, but it should be noted here that city classifiers today, using experience and modern computers, have produced some technically excellent studies and that there are a number of similarities between our classification and others. Indeed, there would almost have to be, since a classification that did much violence to common sense would hardly be useful for policy purposes, or possible to justify.

What, then, is the need for another study? First, our study is up to date, using mainly 1970 Census information. Since many policy relevant parameters of city settings are changing rapidly, such timeliness is essential. Second, most of the earlier studies used a limited set of variables, generally reflecting economic base or demography. In contrast, we have used the experience of early studies to select variables that bear on the currently most important urban problems—including city-suburban differences, local government, and segregation. Most earlier studies classified towns down to a rather small size. We use only large metropolitan areas, thus avoiding factors more relevant to small town classification. In contrast to some studies that seemed to collect data indiscriminately, we selected variables carefully and normalized them by cost of living deflators, population, and other special transformations to better reflect our interest in important *qualitative* differences.

[2] E. L. Thorndike, *Your City,* Harcourt, New York, 1939.

II. THE DATA

Our objective in building the data file was to extract something useful and manageable from voluminous amounts of data available in various sources. A decision was made to restrict the data collection to the 125 urban areas with over 250,000 population;[3] these are pictured in Fig. 1. This lowers costs and focuses attention on phenomena that relate to the larger urban areas, which the Rand Urban Policy Analysis Program was set up to study. In any event, we have included most of the country: The 125 areas contain almost 60 percent of the population of the United States. For reasons of economy, we used mainly data that were already collected and easily available, such as the 1970 U.S. Census of Population and the 1967 City and County Data Book. In some instances, we had to impute missing values by multiple regression.

Variables were selected that bore on some of the major themes of interest of the first set of Rand urban studies. These themes were:

- urban growth and decline
- prosperity and poverty
- race and ethnic minorities
- city-suburban relations
- government and politics

In addition, we included variables giving policy settings and city pathology:

- demography
- geography
- health and crime

The process of building and using the file is on-going; as new important variables are discovered in specific analyses, they are added. A list of the variables and their sources are given in Appendix A.

After the raw data have been collected, they must be carefully normalized to a meaningful form. It is total income or per capita income that is important? Is income itself or its log more appropriate? In Appendix B, we discuss problems of

[3] To ease the collection of data, we used the Standard Metropolitan Statistical Area (SMSA) as our definition of urban area. These are general-purpose units established by the federal government, using counties as building blocks.

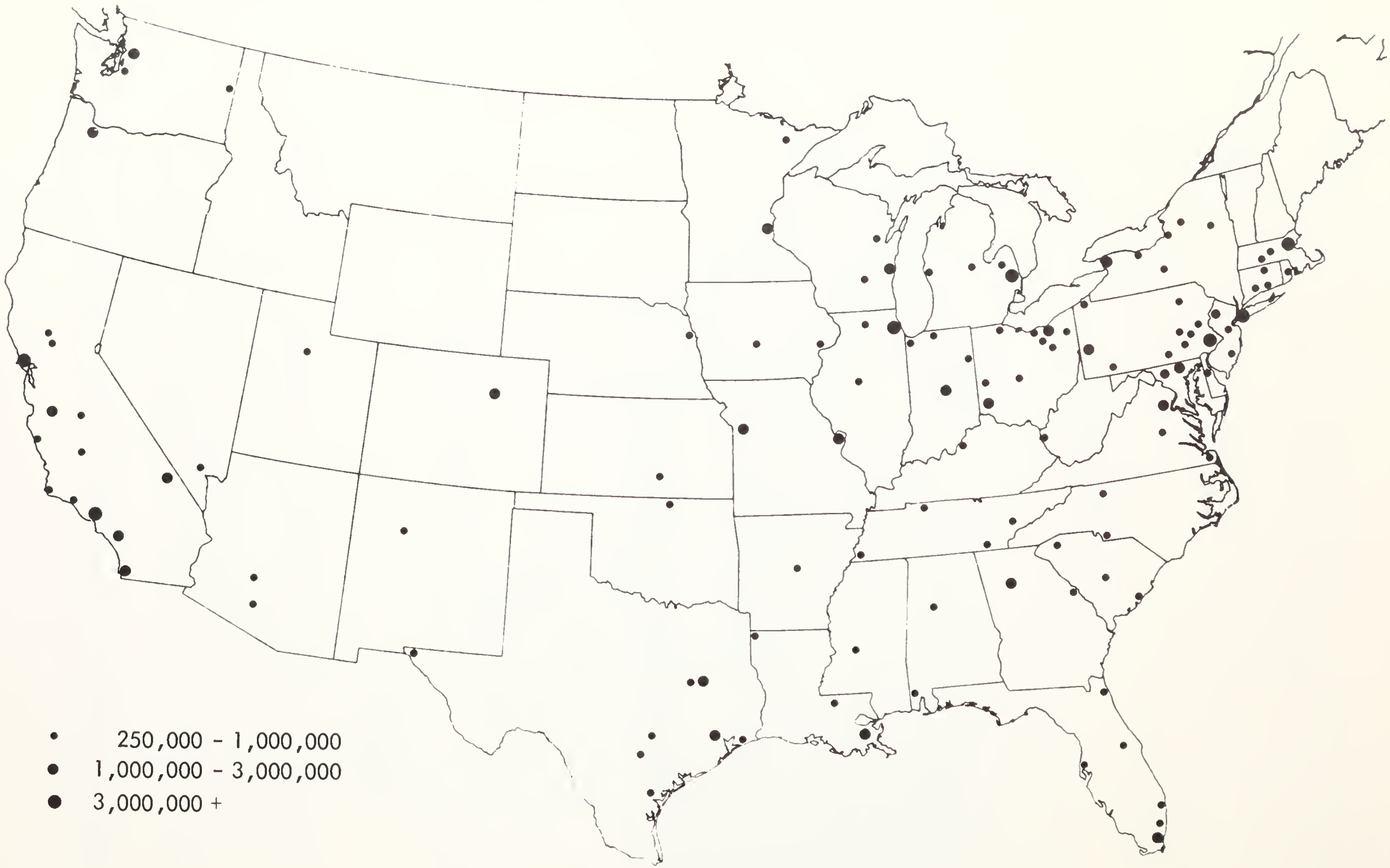

Fig. 1—The location of the 125 largest SMSA's by size

selection and normalization at greater length, but our general approach has been to emphasize qualitative differences. Thus, we have used rates and per capita measures wherever possible.

American urban areas are very diverse. New York has three times as many dentists per capita as El Paso, and 600 times as many robberies per capita as Appleton. Jacksonville has 200 times the area of New York. Tampa has three times as many old people, per capita, as Newport News. These nuggets, gleaned from the tables of the distributions of the 53 variables in Appendix A, are representative. This diversity should not be surprising. Although urban areas are forced by their size and density to have many things in common—policemen, garbage collection, traffic, and so forth—they are nevertheless formed in particular places and with particular people that span the American scene.

III. REPRESENTATION METHODOLOGY

TWO VIEWS OF CITY DIVERSITY

There are two extreme ways of looking at the diversity of cities. One view is that each city is *sui generis,* hence that unique specific local factors are so important nothing can be inferred about one city by studying another. The other is that, for policy purposes, the same model fits all cities—the only difference being different values for certain parameters such as size, income, percent of blacks, and the like. In this model, which underlies statistical cross-sectional analyses of urban phenomena, variations in these parameters have linear additive effects on outcomes. If this view were correct, it would be statistically efficient to study cities with extreme values of the parameters. The remaining cities could be approximated by convex combinations of these extreme cities—for example, Boston = .02 San Jose + .22 Seattle + .02 St. Louis + .02 Little Rock + .4 Philadelphia + .32 Cincinnati.[4] Effects of various policies could be estimated by adding the effects at each city.

We feel that there is a certain truth in both views of city diversity, that although it is possible to learn a lot about cities in general, nonlinearities abound and specific conditions do have a great effect. Thus, instead of the most extreme cities, where indeed some fairly idiosyncratic things may be happening, we are looking for representative ones.

A SYSTEMATIC APPROACH TO REPRESENTATION

It is this diversity of urban areas and the multiplicity of variables that can be used to describe them that confound attempts to select a group of representative areas. Yet research on urban problems, and on policies to ameliorate them, cannot avoid dealing with the problem of representativeness. The cost of a thorough study of even one problem in one city necessitates the selection of a subgroup to represent cities as a whole. Similarly, policy recommendations on a national scale necessitate generalizations in program design and implementation.

It is possible to use an informal approach to representation, with an intuitive

[4] We have computed each city's "best" expression as a combination of the six cities already proposed for study: San Jose, Seattle, St. Louis, Cincinnati, Little Rock, and Philadelphia. Here "best" means lowest squared error for the 53 variables.

balancing of apparently relevant factors, such as size and region. The approach we used, however, is city classification by a method now standard in the field.[5] It involves two steps: The data are organized first by means of factor analysis, and the cities are then classified by means of their factor scores.[6]

Factor analysis is a data description technique that reduces the dimensionality of a set of interrelated variables. It assumes that different phenomena may be part of the same underlying pattern. The technique was first used by psychologists trying to show that two general intelligence "factors" underlay the variety of testable intellectual activities. The term "factor" stuck and is used to connote a synthetic variable that accounts for observed relationships among a set of empirically measured variables.[7]

Without getting into technical details, we would like to show how factor analysis enables us to deal systematically with the city classification. Two terms will be useful:

- *factor loading,* the correlation between one of the empirical variables and a factor.
- *factor score,* the ranking or rating (expressed in standard deviations) of a metropolitan area on the factor that represents a particular pattern of variables.

An example will illustrate how factor analysis reduces the dimensions of variation. In each of the factor analyses performed with various groups of cities, a small but significant factor (stage in life cycle) consistently appeared. It was correlated with a group of variables relating to the age of the population. These included the birth rate, median age, and percent of housing that was crowded. For this factor, presented below in Table 6, the score for El Paso, 3.31, was the highest in the country, and Fort Lauderdale's, −3.27, was the lowest. Within each city, one number represents the whole set of age-related data.

The number of factors used to represent the cities is chosen for explanatory power and intuitive appeal. In our case, the first eight factors explain 64 percent of the total variance of the 53 empirical variables, and (after a standard transformation) each has a clear interpretation. The remaining factors explain less of the variance and do not have a very straighforward interpretation. Each metropolitan area can be represented in an eight-dimensional space by its factor score coordinates. Urban areas that are close together in this space are similar, and those that are far apart are dissimilar. So, we simply divide the areas into "clusters" of neighboring points.[8] The clusters are relatively homogeneous groups of urban areas. The results and models for one metropolitan area in a cluster can generally be expected to carry over to other areas in the same cluster; if not, we will want to discover what important aspects of the urban scene are not covered by our data. By selecting cities from many different clusters, we should be able to cover a large range of American cities.

[5] For an excellent collection of recent work and criticism of the approach, see B. Berry (ed.), *City Classification Handbook: Methods and Applications,* Wiley, New York, 1972.

[6] Appendix B contains a more precise and complete description of factor analysis, the clustering routine, and variable transformations.

[7] Each factor is a linear combination of the original variables.

[8] Our program clusters the points so that the sum of the squared distances from each point to the center of its cluster is minimized. The number of clusters must be selected in advance.

IV. FACTOR ANALYSIS RESULTS

Table 1 presents the factors—the observed underlying patterns in the variables. The factors are named subjectively from the variables with high loadings on the factor. The 53 empirical variables are quite interdependent; eight factors account for 64 percent of their variance. The factors are not absolutely general. Comparison with other studies, presented in Appendix C, shows that the factors depend heavily on the choice of variables and on the group of cities selected. For variables most interesting to us, our factors show what patterns appear in large American cities.

An example may show the advantages and disadvantages of using one number, the "growth" factor score, to replace the variety of growth-related variables.[9] Little Rock is designated an Economic Development Agency growth center and has recently had much commercial growth. However, its score on growth is $-.9$, which puts it in the bottom quarter of cities studied. The low factor score does not mean that Little Rock is not growing in any sense, but that it is quite different from the normal pattern of American "growth" cities: It is losing population, its climate is rather undesirable, there is great inequality in income, and there has been little increase in manufacturing production. Per capita income growth, the only growth variable we collected for which Little Rock is above average nationally, is highly related to poverty and loads mainly on the inequality factor. Even if income growth had loaded on the growth factor, Little Rock would not score high. Since the factor is an *average* of various growth measures, a city must be growing in most of these measures to obtain a high score. The factors are chosen so that they are orthogonal to one another. This orthogonality has important effects. It means that each factor contributes independently to the character of a city, and this makes factors useful in preliminary statistical model building. However, there are disadvantages for interpretation. The whole set must be kept in mind. Education, for example, represents those educational characteristics that remain after we adjust for the "growth" or "poverty" factors. In some ways, the single variable Percent High School Graduates is a better measure of educational status than the educational factor. For our purposes, however, the set of factors locates cities better than a set of corresponding variables that would be significantly intercorrelated.

[9] Too much weight should not be placed on factor names. Population growth correlates .8 with the factor "growth." Thus $(.8)^2$ or .64 of the variance in population growth is predictable from the factor score. The standard deviation of population growth, after adjustment by the growth factor, is $\sqrt{1 - .64} =$ .6 of its former size. This surprisingly small reduction in predictability is brought out in Tables 2-9. Note how the rankings based on factor scores do not lead to consistent rankings based on the key variables in the factor.

Table 1

DIMENSIONS OF THE AMERICAN URBAN SYSTEM IN 1970

Factor Number	Factor Description
1	Nonmanufacturing Economic Base
2	Inequality, Poverty, and Segregation
3	Suburbs and Ghetto Contrasts
4	Growth
5	Stage in Life Cycle
6	Welfare and Unemployment
7	Density
8	Education

Tables 2-9 give the variables with the highest loadings on each factor and the cities with the highest and lowest factor scores. The variables are listed in order of decreasing loadings. "Number" refers to Table A-1 in Appendix A, which has sources and applied transformations given in detail. Loadings are given in percentages. A negative score means low values are associated with the factor. Table A-13 in Appendix A is a list of all the factor scores.

NONMANUFACTURING ECONOMIC BASE

The areas high on this factor are government, recreation, or retirement centers. They are less affected by the business cycle but have more unequal income distributions, possibly reflecting that service jobs are generally lower paying than blue collar manufacturing jobs. Interestingly enough, these areas did much better than the manufacturing cities in getting OEO money. Manufacturing cities are mainly located in the Midwest and Northeast and nonmanufacturing mainly in the sun belt.

INEQUALITY, POVERTY, AND SEGREGATION

This factor picks up different types of variables associated with the deep South. All measures in the variables of local spending and income have been adjusted for cost of living, which is lower for these cities but not low enough to bring the standard of living of their blacks up to the national average. Local segregation, the Barry Goldwater vote, total income, percent black, and the Gini coefficient are measures of conservatism and inequality. Perhaps because legal segregation has ended, there has been much sorting out of the races in these areas by white suburbanization. Thus, the vacancy rate in the cities is high, as is construction employment. There is not the pattern of city apartment-house living, and indeed the SMSAs include some rural areas, so that blacks tend more often to own their residences.

Table 2

NONMANUFACTURING AREAS

Loadings on Constituent Variables: Factor One

Number[a]	Name	Loadings
14	Manufacturing Ratio	-74
25	Gain Value Added (P. Cap.)	-71
24	Change Unemployment	-65
5	New Capital Expenditures	-61
52	Federal Employees	59
42	Anti-Poverty OEO Funds	57
8	Gini Coefficient	56

Manufacturing Areas[b]	Scores	Percent Nonagricultural Employees in Manufacturing
Flint, Michigan	-3.0	46.8
Rockford, Ill.	-2.3	49.1
Wichita, Kansas	-2.1	28.4
Detroit, Mich.	-1.8	37.6
Rochester, N.Y.	-1.7	41.6
Seattle, Wash.	-1.7	24.9
Nonmanufacturing Areas[c]		
Washington, D.C.	2.5	3.8
Tucson, Ariz.	2.2	8.8
Albuquerque, N.M.	2.1	8.6
Salt Lake City, Utah	1.8	16.7
Jacksonville, Fla.	1.7	13.0
Honolulu, Hawaii	1.7	7.4

[a] Number on list in Appendix A.

[b] These areas have the lowest factor scores.

[c] These areas have the highest factor scores.

Table 3

INEQUALITY, POVERTY, AND SEGREGATION

Loadings on Constituent Variables: Number[a]	Name	Factor Two Loading
35	Total Income/Black Income	71
8	Gini Coefficient	69
28	Black Owner Occupied Housing	66
34	Black Family Median Income	-66
47	Local Segregation	64
53	LBJ Vote	-60
44	Infant Mortality	57
45	Cost of Living	-57
3	Construction Employment	54
39	Vacancy Rate	44
46	Income Growth	41
12	Population in 2000	-40

Areas with Great Inequality[b]	Scores	Black Family Median Income
Shreveport, La.	2.8	4635
West Palm Beach, Fla.	2.4	5685
Fort Lauderdale, Fla.	2.4	6676
Jackson, Miss.	2.2	4824
Little Rock, Ark.	1.9	4898
Charleston, S.C.	1.7	5121
Baton Rouge, La.	1.7	5610
Areas of Relative Equality[c]		
San Jose, Calif.	-2.1	10574
Binghamton, N.Y.	-2.1	9558
Lorain, Ohio	-2.0	8614
Jersey City, N.J.	-1.6	7169

[a]Number on list in Appendix A.

[b]These areas have the lowest factor scores.

[c]These areas have the highest factor scores.

SUBURBS AND GHETTO CONTRASTS

When most people talk about the "urban crisis," this factor is probably what they have in mind. These are big cities with rich suburbs on the periphery and black ghettos in the center. In some sense, they are middle-aged cities. In very young western cities, the poorer people live on the outskirts, and the richer live downtown. For such cities as Cleveland, the aging process has hit the center, but the suburbs are still open and new. In the denser and oldest cities, picked out by Factor 7, the suburbs have been urbanized.

RECENT GROWTH

The cities that have been growing over the last few years represent a new type of economic strength—weather and space. Economic growth generally reflects a new type of economic resource. In the United States, the original growth areas were ports with access to agricultural markets, then came railroad and manufacturing centers located by iron and coal sources and, as the country spread, regional market places. Since World War II, weather, beauty, and space have become key considerations—valuable to retirement and to types of new industry that do not need close ties to minerals or older manufacturing centers. The new growth centers have been in the so-called "sun belt." Housing is new and in short supply, so rents are high. The declining areas are in the South and in regions of such declining industries as mining. Although it is not picked up in our data, the same forces are pulling types of light industry that are free to move from the central city to the suburbs.

STAGE IN LIFE CYCLE

Poor rural families are bigger and younger, so on the average blacks and people with Spanish surnames are considerably younger than whites. However, the main reason for difference in this factor is migration: The oldest populations are either the retirement communities of Florida or towns that can't hold on to their young people. The youngest are either rapid growth centers or heavily Spanish speaking, such as in Texas and California. Within most urban areas, the suburbs are considerably younger than the central city, because of their attraction for young (even though white) families. Continuing migration allows areas to specialize in a certain life stage—retirement facilities, say, or suburban family housing. The area stays the same but the inhabitants come and go.

WELFARE AND UNEMPLOYMENT

This is a fairly minor factor reflecting the fact the unemployment is highly tied to welfare and hence to local government expenditures. "Spanish" is in the factor because of its California emphasis; in the 1960s in California, there were many poor people and fairly liberal welfare.

Table 4

SUBURBS AND GHETTO CONTRASTS

Loadings on Constituent Variables: Factor Three		
Number[a]	Name	Loadings
30	Income Suburbs/Central City	80
33	Rent Central City/Suburbs	-72
32	Crowded Housing Central City/Suburbs	70
29	Segregated Suburbs	69
37	Black, Spanish in Central City	59
43	Migration into Central City-Suburbs	-57
18	Median Income	56
40	Robbery	55
1	SMSA Population	53
12	Population in 2000	39

Areas with Great Differences Between City and Suburbs[b]	Scores	Income Suburbs/ Income Central City
Washington, D.C.	3.3	1.35
Newark, N.J.	2.4	1.53
Atlanta, Ga.	2.3	1.27
Detroit, Mich.	2.0	1.21
Baltimore, Md.	1.9	1.20
Wilmington, Del.	1.8	1.33
Cleveland, Ohio	1.8	1.25
Areas with Undifferentiated Suburbs[c]		
Corpus Christi, Texas	-2.2	.95
Tulsa, Oklahoma	-2.1	.94
Appleton, Wisc.	-1.8	.94
Duluth, Minn.	-1.7	.96
Wichita, Kansas	-1.7	.99

[a]Number on list in Appendix A.

[b]These areas have the lowest factor scores.

[c]These areas have the highest factor scores.

Table 5

GROWTH

Loadings on Constituent Variables: Factor Four		
Number[a]	Name	Loadings
4	Change SMSA Population	82
13	Years to 1/2 Size (age)	-76
26	Central City Growth 50-60	74
9	Climate	65
22	Rent	64
21	Black with Both Parents	63
38	Spanish	59
15	Lacking Plumbing	-50
6	Gain Value Added (ratio)	43
41	Burglary	43

Areas of Rapid Growth[b]	Scores	SMSA Population Growth 60-70 (%)
Fort Lauderdale, Fla.	3.4	85.7
Anaheim, Calif.	3.2	101.8
San Jose, Calif.	3.1	65.8
Oxnard, Calif.	2.9	89.0
Santa Barbara, Calif.	2.1	41.2
Las Vegas, Nev.	2.0	115.2
Miami, Fla.	1.7	35.6
San Bernardino, Calif.	1.7	41.2
Declining Areas[c]		
Wilkes-Barre, Pa.	-2.1	-1.3
Duluth, Minn.	-1.8	-4.1
Charleston, S.C.	-1.7	19.4
Johnstown, Pa.	-1.6	-6.4

[a]Number on list in Appendix A.

[b]These areas have the lowest factor scores.

[c]These areas have the highest factor scores.

Table 6

STAGE IN LIFE CYCLE

Loadings on Constituent Variables: Factor Five

Number[a]	Name	Loadings
7	Over 65	-84
2	Birth Rate	79
31	City Pop/SMSA	53
16	Crowding in SMSA	50
50	No. of Govt. Units	-39

Young Family Areas[b]	Scores	Median Age in Central City
El Paso, Texas	3.3	23.2
Honolulu, Hawaii	2.2	28.1
Newport News, Va.	1.7	24.2
Flint, Mich.	1.7	25.2
Areas with Many Old People[c]		
Fort Lauderdale, Fla.	-3.3	39.2
Tampa, Fla.	-2.5	37.8
Wilkes-Barre, Pa.	-2.3	37.6
West Palm Beach, Fla.	-2.3	39.3
Miami, Fla.	-1.8	37.3

[a]Number on list in Appendix A.

[b]These areas have the lowest factor scores.

[c]These areas have the highest factor scores.

Table 7

WELFARE AND UNEMPLOYMENT

Loadings on Constituent Variables: Factor Six		
Number[a]	Name	Loadings
23	Unemployment	71
20	Welfare	70
51	Local Expenditures	56
38	Spanish SMSA	51

Areas with Much Welfare[b]	Scores	Unemployment % 1970
Stockton, Calif.	3.2	8.2
Fresno, Calif.	3.2	6.5
Bakersfield, Calif.	3.1	6.0
Los Angeles, Calif.	2.3	5.8
San Bernardino, Calif.	2.2	5.9
Areas with Low Welfare[c]		
Fort Lauderdale, Fla.	-2.1	2.6
Madison, Wisc.	-1.9	3.1
Appleton, Wisc.	-1.5	4.2
Lancaster, Pa.	-1.5	2.3

[a]Number on list in Appendix A.

[b]These areas have the lowest factor scores.

[c]These areas have the highest factor scores.

DENSITY

The densest cities are the old cities in the Northeast. If we had collected data on the percent of foreign born and percent using public transportation, those variables would also load heavily on this factor.

EDUCATION

This is often used as a measure of social status, complementary to money. Since education is probably the most important determinant of health (income, for example, statistically has a negative effect on health, when education is controlled for),[10] both of the health variables are included here. Suburbs generally show higher on this factor than they do on income differences, since the three main low-education groups—blacks, Spanish, and ethnics—tend to be in the central city.

[10] Michael Grossman, "The Demand for Health," Occasional Paper 119, NBER, Columbia Press, New York, 1972, Chapter VI. His explanation is that poor health may be the result of such typical attributes of higher-income life as anxiety, alcohol, and cigarettes.

Table 8

DENSITY

Loadings on Constituent Variables: Factor Seven

Number[a]	Name	Loadings
27	Density SMSA	91
60	Density Central City	66
1	Population	51
11	One Unit Housing	-50
40	Robbery rates	49

Dense Areas[b]	Scores	SMSA Population per Sq. Mi.
New York, N.Y.	4.4	18,500
Jersey City, N.J.	3.0	13,700
Chicago, Ill.	2.3	7,850
Indianapolis, Ind.	2.1	8,090
Memphis, Tenn.	1.5	3,940
Areas of Low Density[c]		
Oklahoma City, Okla.	-2.1	760
Augusta, Ga.	-2.0	1,040
Mobile, Ala.	-1.9	660
Greenville, S.C.	-1.8	660
Wilmington, Del.	-1.8	1,300
Salt Lake City, Utah	-1.6	1,250

[a]Number on list in Appendix A.

[b]These areas have the lowest factor scores.

[c]These areas have the highest factor scores.

Table 9

EDUCATION

Loadings on Constituent Variables: Factor Eight		
Number[a]	Name	Loadings
17	High School Education	74
19	College Education	59
48	Dentists	54
22	Rent	44
44	Infant Mortality	-31

Areas with Most Educated Population[b]	Scores	% High School Graduates
Seattle, Wash.	2.9	67.8
Minneapolis, Minn.	2.3	66.1
Honolulu, Hawaii	2.3	66.0
Portland, Ore.	2.2	62.9
Salt Lake City, Utah	2.1	68.5
Anaheim, Calif.	1.6	70.5
Boston, Mass.	1.6	64.4
Areas with Least Educated Population[c]		
Jersey City, N.J.	-2.6	36.3
Gary, Ind.	-2.4	50.0
El Paso, Texas	-2.1	51.1
Johnstown, Pa.	-1.7	44.1
Birmingham, Ala.	-1.6	45.4
Greensboro, N.C.	-1.6	42.4

[a]Number on list in Appendix A.

[b]These areas have the lowest factor scores.

[c]These areas have the highest factor scores.

V. CLUSTER ANALYSIS RESULTS

The eight factor scores give a profile of each city. Although these scores locate the city with respect to the national average, further insights can come from distributing the cities into relatively homogeneous groups. To judge how representative a city is, we must know whether there are many cities like it, or whether its profile is unusual. This information will be useful in selecting cities and in weighing contradictory findings from different cities. In addition, the clusters give information on how factors are interrelated. Although they are constrained mathematically to have zero correlation overall, interesting combinative effects appear in the city subgroups. As we shall see, these effects are generally regional—the regions share a climate, history, and economic development that is reflected in the data.

The cities are divided into homogeneous groups as follows: Each city is represented as a point in space, with its eight factor scores as coordinates. The distances between cities have been subjectively weighted so that differences in more important factors have more effect than differences in the other factors.[11] The factor weights are given underneath the factor names in Table 10. A program divides the cities into clusters so that a weighted sum of distances from each city to the center of its cluster is minimized. We present two types of results—the best allocation to ten clusters and a tree of clusters formed by joining the results for different numbers of clusters.

AMERICAN CITIES IN TEN CLUSTERS

Table 10 gives the best ten-cluster results. The top row can be interpreted as follows: The 398.5 is the total weighted squared distance from cities to the center of their clusters. The 1.0, 1.5, and so forth are the subjective weights assigned to the factor. The next line shows that there are 16 cities in the first cluster, that their combined weighted squared distance is 48.95, and that their mean score on the inequality factor is 1.27 standard deviations above average, on the ghetto-suburb factor their mean is .17 and so forth. The next line shows that Columbia, S.C., is closest to the mean for the cluster (only .73 away) and that its score on nonmanufac-

[11] For discussion of how the factor weights are determined, and of the clustering algorithm, see Appendix B.

Table 10

AMERICAN CITIES IN TEN CLUSTERS[a]

			Factor scores[b]							
Cities		Distance to Center	Non-Manufac-turing	Inequality	Ghetto-Suburbs	Growth	Stage of Life Cycle	Welfare	Density	Education
Factor weights			1.0	1.5	1.5	1.5	0.5	0.5	0.7	0.5
Cluster 1:	Average	3.0	.2	1.3	.2	-.4	.5	-.2	-.9	-.5
Columbia	SC	0.7						-		
Nashville	TE	1.2								
Birmingham	AL	1.5							+	-
Chattanooga	TN	1.6	-			-	-			
Mobile	AL	1.6						+	-	-
Charlotte	NC	2.2				+		-	+	
Baton Rouge	LA	2.5					+		+	++
Jacksonville	FL	2.6	++							
Greenville	SC	3.0	-					-	-	-
Jackson	MS	3.6		+	-					
Greensboro	NC	3.7		-	-	+		-	+	-
New Orleans	LA	3.7	+		+			++		
Charleston	SC	4.0	+			--	+			
Beaumont	TX	4.9	--			+			-	
Augusta	GA	5.6			++	-		-	-	
Oklahoma City	OK	5.9	+	--					--	++
Cluster 2:	Average	2.5	.7	-.4	-.5	.8	-.4	2.3	-.5	-.5
San Diego	CA	1.2						--		
Stockton	CA	1.4				-		+		
Salinas	CA	2.1			-	+				
Bakersfield	CA	2.4						+	-	
San Bernardino	CA	2.7				+	-		-	
Fresno	CA	3.2			-	-		+		
Los Angeles	CA	3.5	-		+				++	
Sacramento	CA	3.5	+	-	+			-		+
Cluster 3:	Average	2.4	-1.3	-.5	-.2	0	.4	-.2	0	.1
South Bend	IN	0.5								
Toledo	OH	0.7							+	
Akron	OH	0.8						-		
Fort Wayne	IN	0.8					+	-		
Grand Rapids	MI	1.0								
Lansing	MI	1.0								+
Bridgeport	CT	1.2			+		-			
Davenport	IA	1.2			-				-	
Canton	OH	1.4						-		-

Table 10 (Cont.)

AMERICAN CITIES IN TEN CLUSTERS

Cities		Distance to Center	Factor scores[b]							
			Non-Manufac-turing	Inequality	Ghetto-Suburbs	Growth	Stage in Life Cycle	Welfare	Density	Education
Peoria	IL	1.5	+							
Youngstown	OH	1.5							-	-
Rochester	NY	2.8			+		-			+
Rockford	IL	2.9	-			+				-
Tacoma	WA	3.2	+					+	--	+
Indianapolis	IN	4.5	+					-	++	
Flint	MI	6.1	--		+		++	++		
Seattle	WA	6.1						++		++
Lorain-Elyria	OH	6.4	+	--			+	-	-	--
Cluster 4: Average		3.6	-.1	1.1	-1.5	-.8	.2	.2	.7	0
Knoxville	TE	1.0								
Memphis	TE	1.9							+	-
Tulsa	OK	2.0			-				+	+
Little Rock	AR	2.2		+						
Omaha	NE	2.7		-				-		+
Huntington	WV	2.8		-					-	-
Shreveport	LA	4.6		++						
Wichita	KA	5.8	--				+	+		+
Corpus Christi	TX	6.4	+		-	++	++	+		--
Wilkes-Barre	PA	6.6		-		--	--		-	
Cluster 5: Average		4.7	.2	1.4	.3	1.5	-1.1	-.5	0	-.1
Dallas	TX	2.2					++			+
Tampa	FL	2.3	+				--			-
Houston	TX	2.5					++			
Phoenix	AR	2.9			-		+	+		+
Orlando	FL	3.4	+	-					-	
West Palm Beach	FL	4.5		+			--		-	
Fort Worth	TX	5.2	--				++			
Miami	FL	5.4	+		++		-	++	+	-
Fort Lauderdale	FL	13.3		+	-	++	--	--	++	
Cluster 6: Average		3.0	.1	-1.0	-.7	-.7	-.7	-.6	-.6	0
Worchester	MA	0.5								
Erie	PA	1.2	-							
Albany	NY	1.4	+		+					
Reading	PA	1.6	-				-		+	-
Utica-Rome	NY	1.6							-	
Allentown	PA	1.8					-	-		-

Table 10 (Cont.)

AMERICAN CITIES IN TEN CLUSTERS

Cities		Distance to Center	Factor scores[b] Non-Manufac-turing	Inequality	Ghetto-Suburbs	Growth	Stage in Life Cycle	Welfare	Density	Education
York	PA	1.9	-							-
Lancaster	PA	2.3			+			-		-
Des Moines	IA	2.5	+				+			+
Spokane	WA	3.2			-			+		++
Johnstown	PA	3.4				-	-	+		--
Binghamton	NY	3.9	-	-		+		+		-
Appleton	WI	4.9	-		-	+		-	+	
Duluth	MN	6.4	++		-	-			-	+
Salt Lake City	UT	7.8	++				++	+	-	++
Cluster 7:	Average	3.5	-.1	-.2	1.8	-.1	.2	0	.4	-.1
Cleveland	OH	0.3								
Baltimore	MD	1.0							+	-
Trenton	NJ	1.1								
Hartford	CT	1.6					-		-	+
Patterson	NJ	2.2		-			--			
Dayton	OH	2.3	-					-	-	
Richmond	VA	2.3		+				-		
Newark	NJ	2.4			+		-	+	+	
Chicago	IL	2.9							++	
St Louis	MO	3.7		+		-		+		
Atlanta	GA	3.9		+					--	+
Wilmington	DE	4.3							--	
Detroit	MI	5.1	--				+	++	+	
Gary	IN	6.5	-	-	-	+	++			--
Washington	DC	12.4	++		++		+	-		++
Cluster 8:	Average	2.8	-.3	-.8	-.3	2.7	.2	.5	-.3	.9
Oxnard-Ventura	CA	0.9								-
Las Vegas	NV	2.4	+	+		-	+			
Santa Barbara	CA	2.5	+				-		-	
San Jose	CA	3.9		--			+		+	-
Anaheim	CA	4.3	-	+			-			+
Cluster 9:	Average	3.8	1.5	-.3	-.5	.6	1.3	-.7	.2	-.1
San Antonio	TX	2.4						+		--
Austin	TX	3.1		+			-	-	+	+
Albuquerque	NM	3.5	+	-	-			++		
Tucson	AZ	3.5	+			+	--			
Madison	WI	3.6		-			-	-		++

Table 10 (Cont.)

AMERICAN CITIES IN TEN CLUSTERS

Cities		Distance to Center	Factor scores[b]							
			Non-Manufac-turing	Inequality	Chetto-Suburbs	Growth	Stage in Life Cycle	Welfare	Density	Education
Norfolk	VA	3.7			++	-				-
Honolulu	HA	3.8					+			++
Newport News	VA	3.9	-		+				--	
El Paso	TX	6.5		-	-		++	+		--
Cluster 10:	Average	3.0	.1	-.4	.3	-.6	-.4	.2	.6	.4
Pittsburg	PA	0.8					-			
Buffalo	NY	0.8								
Syracuse	NY	0.9							-	
Milwaukee	WI	1.1	-					-		
Springfield	MA	1.2							-	
Providence	RI	1.3				-			-	
Cincinnati	OH	1.3							-	-
New Haven	CT	1.4	-						-	
Philadelphia	PA	1.7			+				+	-
Columbus	OH	1.8					+	-		
Louisville	KY	2.1		+			+			
Kansas City	MO	2.2		+					-	
Boston	MA	2.3	+				-			++
Minneapolis	MN	2.6						-		++
Denver	CO	2.6	+			+				+
Harrisburg	PA	3.1	+				-	-	--	
San Francisco	CA	3.6	+					++		+
Portland	OR	3.7			-			+	-	++
Jersey City	NJ	11.5		--					++	--
New York	NY	12.6	++					+	++	-

[a]The symbols in the table show the relative position of the cities in the cluster:

- \- .6 to 1.19 standard deviations below the cluster average
- -- 1.2 or more standard deviations below the cluster average
- \+ .6 to 1.19 standard deviations above the cluster average
- ++ 1.2 or more standard deviations above the cluster average.

[b]Given in standard deviations from national average.

turing is (.23 + .03) = .26. The numbers are relative to the mean for the cluster. Thus Columbia's score on the inequality factor is 1.54.

At the bottom of the list, the cities may be less representative of their cluster or, as is Fort Lauderdale in cluster 5, extreme cases of what the cluster represents. By looking at the residuals, we get an idea of how such cities differ from the main body of the cluster.[12] Thus, in cluster 10, New York and Jersey City are much more dense than the rest of these dense cities. If more clusters were allowed they would break off into their own two-city cluster. As it is, every city, no matter how special, must go somewhere.

We shall discuss the satisfactoriness of the clustering after we describe the ten clusters. Cluster 1 and cluster 4 are southern cities. Those in cluster 1 are older and less dense with differentiated suburbs. Cluster 2 contains the less prosperous California cities. They are very high on unemployment and welfare. Their growth rate is above average, but it is nowhere near as fast as cities in cluster 8, the California boom towns. Many of these cluster 2 cities are farm centers, with less education and manufacturing than cities in cluster 8. Cluster 3 contains the manufacturing cities of the midwest. In cluster 5 are the Texas and Florida growth cities. They are poor and black and have older residents than the average city. Cluster 6 consists of the declining, white, smaller Northeastern cities. In cluster 7 are the big cities with large black ghettos ringed by prosperous white suburbs. The other big cities, in cluster 10, are somewhat denser, and have suburbs that are either smaller or more like the central cities. Cluster 9 is made up of the nonmanufacturing cities. They are young, growing, and mainly in the southwest.

The ten-cluster solution presented here was the most satisfactory one produced. It contains few incongruities as seen by urban experts. Some incongruities are unavoidable since there are cities that are unique in ways in which we have not been able to collect data. We would hope that many of these would be on the edges of the clusters. The clusters are in no sense uniquely determined; many transitions are possible with little effect on the total score. For example, Härrisburg is in most ways more like the cities in cluster 5 than those in cluster 9; the main difference is in suburb differentiation. In different runs, the types of clusters generally stay the same, but a few cities in the clusters may change.

One problem with the factor-cluster method is that it is an average classification.[13] Although it may have some relevance to many urban problems, it does not classify areas exactly according to any one problem. This is what we want in our selection of representatives; but if we were trying to study one particular problem, we might return to the data file to select cities precisely on the variables affecting that problem.

A TREE OF CLUSTERS

The clustering analysis was repeated with 4, 6, 8, 10, and 14 clusters. The result is the somewhat hierarchical clustering shown in Figure 2. Each cluster is identified

[12] We have computed scatter plots of the cities on pairs of factors, which show more clearly the relation of cities to clusters.

[13] C. A. Moser and W. Scott, *British Towns,* Oliver and Boyd, Edinburgh, 1961.

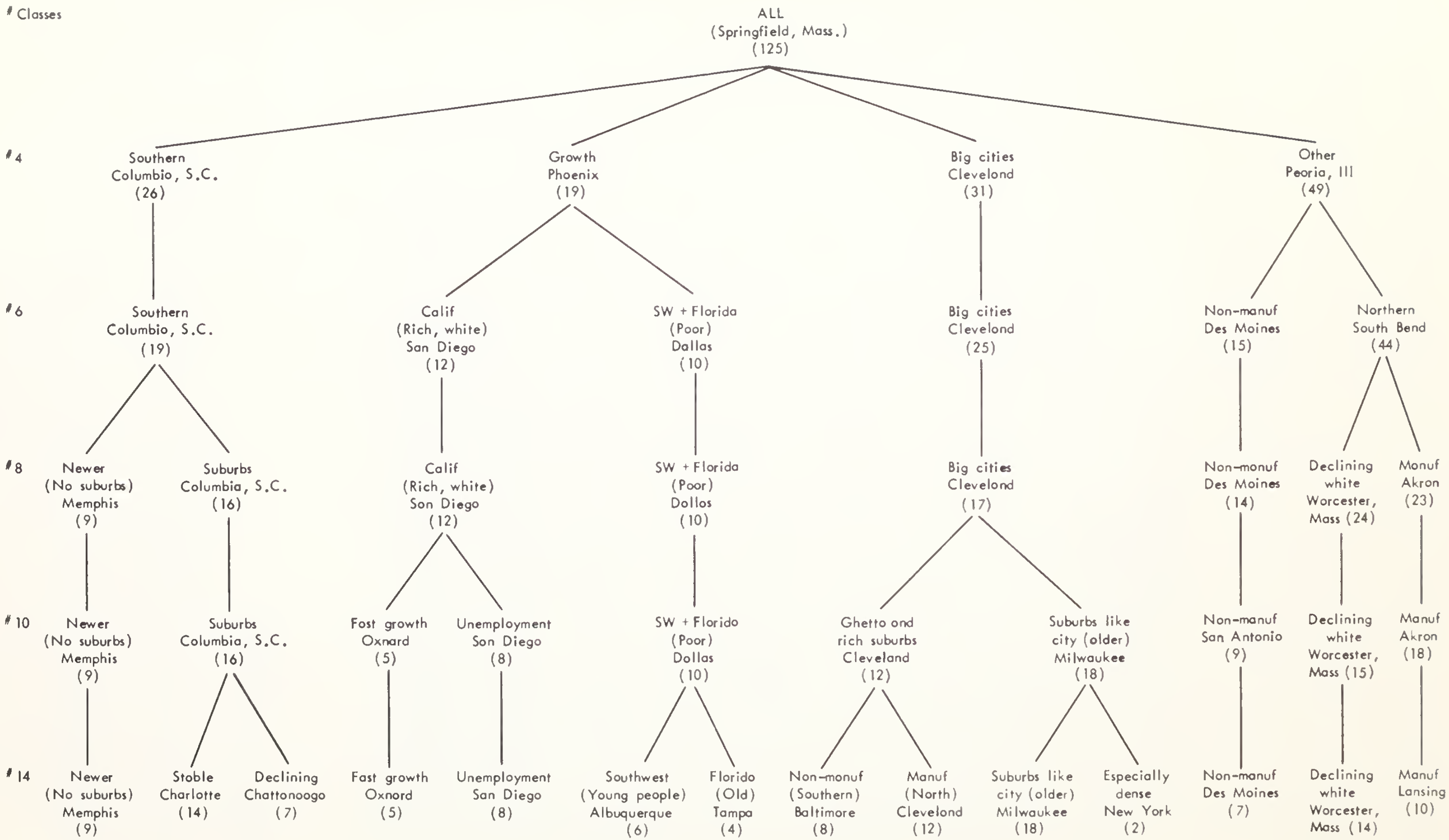

Fig. 2—Tree of classes of 125 largest SMSA's clusters identified by salient characteristics and most representative city (number of cities in cluster in parentheses)

by its salient characteristic, its most representative city, and the number of cities in the cluster. It is not strictly hierarchical, since the new, more specialized clusters that form in lower levels may pick up cities from other more general clusters. For example, when the big cities with ghettos on the #10 line split into northern manufacturing and southern nonmanufacturing cities, they add Akron, Bridgeport, and Dayton from the manufacturing cluster, and Norfolk from the nonmanufacturing cluster. The tree clearly brings out the *regional basis of city differentiation.* The cities are clustered in terms of manufacturing or nonmanufacturing, inequality or affluence, young or old, big or small; but most are just as accurately described as Florida, California, or New England clusters.

VI. CONCLUSIONS

We have tried to make a simple characterization of the 125 U.S. cities with populations greater than 250,000. The first step was to determine a set of variables that measures their diversity; we focused on a total of 53, covering growth and decline of areas, social and racial inequality, government and political structure, and city-suburban relationships. Clearly, any such set of variables contains a good deal of redundancy, so the next step was to eliminate it. We used factor analysis to find eight variables that were linear combinations of the original 53 and account in some sense for most of their variability. Thus, each city was represented as a point in eight-dimensional Euclidean space. To determine which of these points are close, we applied a clustering algorithm in generating ten fairly homogeneous groups. These groups are mainly regional, since the common history and economic development of regions are reflected in the data.

Alford argues that a major problem of classifications is not that they are useless, but in practice no one uses them.[14] Our work has already been used in the selection of these future sites for study: Little Rock, Cincinnati, and Philadelphia. We hope that a variety of useful models, such as the one presented in Appendix D, can be developed and tested. It seems to be a very cost-effective investigatory procedure.

[14] R. Alford, "Critical Evaluation of the Principles of City Classification," in B. Berry (ed.), *City Classification Handbook.*

Appendix A

DATA SOURCES AND REMARKS

Table A-1

LIST OF VARIABLES[a]

Number of Variable	Description of Variable	Source/Table	
1	Population (log people)	SA	3
2	Birth rate, 1968 (per 1000 population)	VS	
3	Construction employment (% non-agr. empl.)	SA	71
4	Population change, 1960-70 (log (70 pop/60))[b]	SA	5
5	New capital expenditures 1963 (log $ per capita)	CCD3	70
6	Gain in value added (ratio 1967/1963)	SA	192, 197
7	Over 65 years old (% pop)	SA	20, 43
8	Gini coefficient (a large value for inequality)	USC	89
9	Climate index (subjective desirability regressed on temp. and rain)[b]	CCD4	341
10	Area of central city (log sq miles)	SA	11
11	One unit structures (% year round units)	SA	96
12	Population of year 2000 Metropolis (log people)[b]		
13	Years since city was half of present size (log)[b]		
14	Manufacturing ratio (% non-agr. empl.)	SA	67
15	Lacking plumbing, housing (% occupied units)	SA	91
16	Crowded housing (% households)	SA	92
17	High school graduates (% pop)	USC	83
18	Family median income ($1000)	USC	89
19	College graduates (% pop)	USC	83
20	Welfare, 1971 (AFDC as % pop)	SA	165
21	Black children living with both parents (% black kids)	USC	90
22	Rent (monthly median in $)	SA	108, 118
23	Unemployment (% work force)	SA	80
24	Unemployment increase 1969-70 (% work force)	SA	78
25	Gain in value added per capita (1967/1963 in $1000)	SA	192, 197
26	Population change 1950-60 in central cities (log (60 pop/50))[b]	SA	14
27	Density of population (log people per sq mile)	SA	11, 33
28	Black owner occupied housing (% total occupied units)	SA	94
29	Segregation of suburbs from city (% complete segregation)[b]	SA	36, 40
30	Income ratio, suburbs/central city (median family incomes)	USC	89
31	City population (% SMSA pop)[b]		
32	Crowding (% central cities-% suburbs)	SA	92, 103
33	Rent (monthly median, central cities - suburbs)	SA	108, 118
34	Black family median income ($1000)	USC	94
35	Income (white/black, family median)	USC	89, 94
36	Nonwhite infant mortality - white (% of births, 1967)[b]	VS	
37	Black and Spanish in main city (% of pop)	USC	81
38	Spanish (% of pop)	USC	81
39	Vacancy rate in central cities housing (% units)	SA	99, 100
40	Robbery per capita (per 1000)	SA	180
41	Burglary per capita (per 1000)	SA	182

Table A-1 (Cont.)

LIST OF VARIABLES

Number of Variable	Description of Variable	Source/Table
42	OEO antipoverty funds allocated ($ per capita)	SA 160
43	Relative migration, 1960-70 (main city rate - suburb)[b]	
44	Infant mortality, 1967 (infant deaths per births)	VS
45	Cost of living, 1969 (expenses for intermediate family of 4/$9000)[b]	
46	Income growth, 1959-69 (annual rate of per capita income growth)	SA 5, 64
47	Tauber's segregation index, main city 1960 (computed block by block)[b]	
48	Dentists (# per 100,000 pop. in 1969)	SA 58
49	Type of government (weak mayor=0, stronger=1.2, comm.=3, manager=4)	MY
50	# governmental units, 1967 (normalized by pop, log)[b]	SA 140
51	Local direct general expenditure (log $ per capita)	SA 146
52	Federal government employees, 1969 (% pop)	SA 152
53	LBJ 1964 vote (% total presidential vote)[b]	SA 133

[a]Unless stated otherwise, data are for the SMSA in 1970.

SOURCES: MY = *Municipal Yearbook*, ICMA, Chicago, 1972. SA = *Statistical Abstract*, 1971, pp. 830-889. CCD3 = *City and County Data Book*, 1967, Section 3. CCD4 = *City and County Data Book*, 1967, Section 4. USC = Census of Population of General Social and Economic Characteristics, 1970. VS = *Vital Statistics of the U.S.*, 1967.

[b]Variable

4. Based on Area of 1970 SMSA.

9. Ten people were asked to rank ten major cities on a scale of 1-5 for desirability of climate. The average ranking was regressed against summer and winter mean noon temperature and annual rainfall. Desirability = .077 Winter - .030 Summer - .035 Rainfall.

12. Jerome P. Pickard, *U.S. Metropolitan Growth and Expansion, 1970-2000 with Population Projections*, Urban Land Institute, Washington, 1971, Tables III-6 through III-8, with the low census-E projections of birth rate. There are three Megapolitan areas--Atlantic Seaboard, Lower Great Lakes, and California--and other smaller areas.

13. Generated at Rand from the 1950 *Statistical Abstract*, the *Encyclopedia Britannica* and some guesswork.

26. Data based on 1970 areas of cities.

29. Percent of whites who must move to make percent whites equal in city and suburb, as a fraction of the percent who must move to integrate the city if segregation were total.

31. 1971 *Statistical Abstract*, p. 21.

36. Since some cities have very small nonwhite populations, an experimental Bayes technique was used. The corrected infant mortality rate was (Nonwhite Infant Deaths in 1967 + 9)/(Nonwhite Births in 1967 + 250). It is essential that 9/250 = .036, the national nonwhite ratio of infant deaths to births.

43. Data from the working file of P. A. Morrison, Rand.

45. The 1970 *Statistical Abstract*, p. 346, gave the 1967 estimated costs of living for an urban family of four in 34 of our 125 metropolitan areas. For the other 91, we used the regressed estimate, Cost of Living = 6126 + .536 (Per Capita Income) + 27.3 (Latitude) - 7.18 (% Minority) - 309 (if in South). This had an R^2 of .76 for the 34 cities.

47. Taken from K. E. Taeuber and A. F. Taeuber, *Negroes in Cities*, Aldine, Chicago, 1965, p. 32. The index computes segregation in 1960 as in variable 29, but on the basis of census blocks, rather than just city and suburbs.

50. Normalized by dividing by the square root of population. This is supposed to allow for a naturally greater number of governments where there is a greater number of communities coming together.

53. A measure of conservatism. In this election, there were few minor party votes.

The diversity of American urban areas is shown in Tables A-2 through A-10, tables of empirical variables. Five evenly spaced points on the distribution of 125 areas are given for each variable: The lowest, 32nd lowest (25 percent), 63rd lowest (median), 94th lowest (75 percent), and highest values of the variable. The city with that particular value is also shown. These order statistics are preferred to the mean and standard deviation because of their insensitivity to scaling and extreme values. The numbers speak for themselves.

Table A-11 shows the skewness and outliers in the data. In this table +, ++, and +++ represent high outliers of the variables more than one, two, or three standard deviations above average; –, – –, and – – – indicate low outliers. It is apparent from the table that certain cities, such as Fort Lauderdale or New York, are extremely different from the average on many variables.

Table A-12 gives the correlation coefficients between each pair of variables. These coefficients measure the observed simple linear relationships in the data. These relations may be accidental, or they may reflect the implications of true cause and effect. A positive coefficient means that variables are directly related, and a negative value indicates an inverse relationship. The closer the correlation is to ± 1, the closer one variable is to being a linear transformation of the other. To interpret such relationships, we must control for the influence of other variables, as our later multivariate analysis does. It should be noted that the coefficients measure only the degree of linear relationship; significant nonlinear relationships may exist but still yield small coefficients.

Table A-2

URBAN GROWTH AND DECLINE[a]

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Construction Employment (percent non-farm employment)	3	2.7 JC	4.1 SBD	4.8 SF	5.8 DLS	13.6 FL
Population Change 1960-70 (%)	4	-6.4 JHN	9.5 SPD	16.1 GRO	24.5 SLC	115.2 LV
New Capital Expenditures (1963 $ per capita)	5	8 ASN	35 SLC	60 CNI	79 DYN	416 HNN
Ratio Value Added 1967/1963	6	.98 LRN	1.27 CLD	1.37 KC	1.45 PRD	2.51 FL
Gain in Value Added per capita ($ 1967/1963)	25	-184 LRN	160 BLE	311 STN	449 TLA	1,447 BMT
Population Change in Central Cities 1950-60 (%)	26	-15 WB	-2 SYE	14 SBD	41 ASN	368 TCN

[a]See Table A-11 for city abbreviations.

Table A-3

INCOME AND EDUCATION

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Gini Coefficient (Families, SMSA)	8	.286 LRN	.319 ANM	.337 SL	.361 LA	.424 WPB
Family Median Income[a]	18	8,035 ALN	9,585 SLS	10,262 CNN	10,749 LV	12,556 WSN
Per capita Income Growth (1959-60, %)	46	26.5 OXD	61 FRO	68 SPE	75 MBE	119 AGA
Cost of Living (Family of four, $)	45	7,820 CC	8,440 SB	8,960 YNN	9,180 TRN	10,330 HNU
High School Graduate (%)	17	36.3 JC	50.7 KNE	54.2 SBD	60.1 MBE	71.3 SBA
College Graduates	19	3.3 MBE	9.3 PRA	10.8 BRT	13.2 PTN	23.4 WSN

[a]Adjusted by cost of living.

Table A-4

POVERTY: HOUSING AND UNEMPLOYMENT

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Lacking Plumbing Housing (%)	15	.6 ANM	2.4 LRN	3.1 MLE	4.2 UTA	11.8 CHN
Crowded Housing (% Households)	16	3.4 RDG	6.1 ANM	7.1 FL	9.2 CLA	19.9 HNU
Rent[a] (Monthly Median)	22	53 JHN	78 BFO	92 ALE	107 FLT	142 SJ
Unemployment (% 1970)	23	2.1 RCD	3.5 OC	4.2 DLH	5.4 BFO	9.5 SEE
Unemployment Increase 1969-70	24	0 DM	.7 OMA	1.1 BSN	1.3 CLD	5.5 SEE
Welfare (AFDC as % Population 1970)	20	1 APN	3.5 ASN	5.5 HNU	6.5 CLD	14 BNN

[a]Adjusted for cost of living.

Table A-5

RACE AND RACIAL DIFFERENCES

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
% Black Children Living with Both Parents	21	29 WB	55 MDN	57 NO	60 UTA	81 HNU
Black Owner Occupied Housing (% All Housing)	28	0 APN	1 HNN	3 HRG	7 CLD	16 JCN
Black Family Median Income	34	4,635 SHT	6,177 SA	6,779 RDG	7,329 SBD	10,574 SJ
Black, Spanish in Main City (%)	37	1 APN	14 YRK	25 LR	35 HRD	73 WSN
Spanish in SMSA (%)	38	0 ALN	1 NN	1 SPD	6 DM	25 MMI
Median Income/ Black Median (Families)	35	1.03 BNN	1.43 PHA	1.51 DVT	1.61 SYE	1.94 SAT
Adjusted Nonwhite Infant Mortality minus White (% Births)	36	-.3 CC	1.1 LV	1.5 DYN	2.0 RDG	3.1 BNN

Table A-6

SUBURBAN AND CITY CONTRASTS

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Income Ratio Suburb/City[a]	30	.891 BKD	1.008 DVT	1.063 GR	1.142 RDG	1.532 NWK
Monthly Rent Suburb-City ($)	33	-30 TLA	-4 JHN	8 CNN	22 BLE	54 DTT
Crowded Housing % City-Suburb	32	-2 ALE	-.3 RDG	0 WB	.3 JC	3.5 NWK
City Population (% SMSA Population)	31	9 SB	27 WSN	38 DLH	51 HNU	100 JCE
Taeuber's Index of Segregation 1960 (100=complete)	47	72 SJ	80 TCA	87 ERE	92 FWE	98 FL
Vacancy Rate City Housing (% Units)	39	0 JHN	5 LA	6 FRO	7 FW	10 FL
Relative Migration (Total City Rates) (1960-70)	43	32 AGA	15 BRM	7 SPE	-7 OXD	-46 KNE

[a]Jacksonville, whose central city is its SMSA, is not considered to have suburbs in our data. It has been given the average ratio, where appropriate.

Table A-7

GOVERNMENT AND POLITICS[a]

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Governmental Units in SMSA, 1967 (Ranks adjusted for population)	50	4 HNU	37 CHN	75 WPB	137 WCA	704 PTG
OEO Anti-Poverty Funds ($ per capita)	42	1.5 ANM	5.6 OXD	7.7 TLO	10.8 TMA	44 JCN
Local Direct General Expenditure ($ per capita)	51	91 HNU	203 BR	225 NSE	270 TCN	459 STN
LBJ Presidential Vote, 1964 (%)	53	11 JCN	53 CHE	62 SLS	67 FLT	82 PRE
Federal Government Employees (1969) (% Population)	52	.3 GRY	.6 ANM	1.0 CLD	1.8 PHA	10.9 WSN

[a]Two other variables are discrete: 29 Areas are considered to have "Congressional Power," by virtue of being represented by a congressman or senator who heads a major committee; 43 of the central cities are led by a manager, 15 are governed by a commission, and the rest have mayors.

Table A-8

URBAN DEMOGRAPHY

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Population, SMSA, 1970	1	250,000 SLS	320,000 LNR	541,000 NSE	1,013,000 TMA	11,529,000 NY
Population Density (SMSA) (People per square mile)	27	650 DLH	1,880 DLS	2,460 SBD	3,300 PRA	18,540 NY
Birth Rate (1968 per 1000)	2	13.6 FL	16.9 NH	18 BKD	19.1 HSN	28.9 EP
Over 65 (%)	7	7 NN	8.5 DYN	9 SBA	10 BFO	20 TMA

Table A-9

URBAN GEOGRAPHY

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Climate Index	9	-23 DLH	-14 LNG	-10 MMS	-1 AGA	24 HNU
Area of Central City (square miles)	10	4.5 YRK	24 KNE	44 SPD	71 TCN	769 JCE
One Unit Structures (% Units)	11	13.5 JC	63.3 MNS	72.7 SLC	77.6 APN	85.4 BMT
Expected Population of Year 2000 Megalopolis	12	220,000 DLH	3,600,000 MNS	35,000,000 LA	52,000,000 CHO	57,000,000 NY
Age of City (years since one half present population)	13	8 SJ	26 SPE	50 LR	70 DLH	180 CHN
Manufacturing Ratio (% non-farm employees)	14	4 WSN	18 TMA	27 LNG	35 UTA	49 RKD

Table A-10

HEALTH AND CRIME

Variable Name	Variable Number	Lowest Value	25 Percent	Median	75 Percent	Highest Value
Infant Mortality 1967 Deaths/Births (%)	44	1.2 YRK	1.9 SF	2.1 PTG	2.9 MMI	3.0 CLA
Dentists (per million, 1969)	48	326 EP	465 CNN	536 HRG	642 OMA	957 NY
Robbery Rate (per 1000 population, 1969)	40	.1 APN	.8 YRK	1.23 CNN	1.83 JC	6.65 NY
Burglary Rate (per 1000 population, 1969)	41	2.75 WB	8.6 BFO	12.3 NH	15 BKD	22.7 JCE

Table A-11

DATA SUMMARY

TAG	-----CITY------		POPUL 1	BIRTH 2	CONEM 3	POPCH 4	CAPEX 5	VARAT 6	OLD 7	GINI 8	CLIM 9	DENSC 10	1 UNIT 11	POP 20 12	AGECI 13	MANUF 14	-PLUM 15	CROWD 16	HIGHS 17	INCOM 18	COLLG 19	WELF 20	BLFAM 21	RENT 22	UNEMP 23	UNINC 24	VAINC 25
AKN	AKRON	OH			-		+			-				+		+		-		+		-					+
ALY	ALBANY	NY							+		-	-	-	+	+			-				-		-			
ALE	ALBUQUERQUE	NM	-		+		--	-	-	+	+		+	-	-	-		+	+	-	+	+					-
ALN	ALLENTOWN	PA		-	-		+		+	-	-	-		+		+		-	-	--	--	-		-	-		
ANM	ANAHEIM	CA	+			+++		+	-	-	++				-		--		++	++	+		+	++	+	+	
APN	APPLETON-OSKS	WI	-		-		+			-	-	-	+	+		+						--	+				
ATA	ATLANTA	GA	+	+		+			-			+		-						+	+			+	-		
AGA	AUGUSTA	GA	-	+					-	+		-	+	-	+		+	+	-	-		+		-			
ASN	AUSTIN	TX	-		+	+	--		-	+	+				-	-			+		++	-		+	-	-	-
BKD	BAKERSFIELD	CA	-					-			+	-	+		-	-	-	+		-	-	+	-	-	+		-
BLE	BALTIMORE	MD	+					-						+	+				-	+		+	-				-
BR	BATON ROUGE	LA	-	+	+++		+	-	-	+			+		-	-		+			+				+		
BMT	BEAUMONT	TX	-		+	-	++	+					+					+	-		-	-	+	-		-	+++
BNN	BINGHAMTON	NY	-			-		+		-	-	-		+	+	+		-				+++	+				+
BRM	BIRMINGHAM	AL				-				+			+	+			++		-	-	-			-		-	
BSN	BOSTON	MS	++		-				+		-		--	+	+			-	+		+		-				
BRT	BRIDGEPORT	CT		-						-	-		-	+	+	+		-		+		-		+	+	+	+
BFO	BUFFALO	NY	+		-	-	+			-	-		-	+	+		-	-					-	-			
CNN	CANTON	OH			-		+			-		-	+	+		+					-	-	+	-			+
CHN	CHARLESTON	SC	-	++	+				-	+		--		-	++	-	+++	+	-	-		-					-
CHE	CHARLOTTE	NC		+	+				-						-							-		-	-	-	
CHA	CHATTANOOGA	TN	-			-		+		+				-		+	+		-	-			-	-			+
CHO	CHICAGO	IL	++		-						-	+	--	+	+					+				+			
CNI	CINCINNATI	OH	+										-	+	+				-								
CLD	CLEVELAND	OH	+			-	+						-	+	+		-	-		+				+			
CLA	COLUMBIA	SC	-		+				-	+		-	+	-		-	++			-	+	-		-			-
CLS	COLUMBUS	OH						-				+		+				-	+		+				-	-	-
CC	CORPUS CHRSTI	TX	-	+	+	-	--	++	-	+	+		+	--	-	-	+	++	-	-			+	-			
DLS	DALLAS	TX	+	+		+			-	+		+			-					+	+			+	-	+	
DVT	DAVENPORT-RI+	IA	-				+	+		-	-			-							-	-				-	+
DYN	DAYTON	OH			-					-				+		+				+		-			-		+
DNR	DENVER	CO	+			+		-						-		-		-	+		+			+	-	-	-
DM	DES MOINES	IA	-			-					-			--					+						-	-	
DTT	DETROIT	MI	++		-		+					+		+		+	-			++				+	+	++	
DLH	DULUTH	MN	-	-	-	-	-	-	+	-	-			--	+	-	++			-		-	-	-			-
EP	EL PASO	TX	-	+++			-	+	-	+	+	+		-	-		++	+++		-		-	++				
ERE	ERIE	PA	-		-	-				-	-	-		+		+				-	-			-			+
FLT	FLINT	MI		+	-		+		-	-	-		+	+		+				+	-	+	+	+	++	+++	+
FL	FORT LAUDRDLE	FL		--	+++	+++	--	+++	+++	++	+	-	-		-	-	--					-	+	++	-		
FWE	FORT WAYNE	IN	-	+						-	-			+		+	-					--				+	+
FW	FORT WORTH	TX				+		++				+	+				-			+		-				+	+
FRO	FRESNO	CA			-			-		+	+		+		-	-	-	+		-		+++	-		+	-	-
GRY	GARY	IN					++	-	-	-	-			+	-	+		+	-	+	-		+				-
GR	GRAND RAPIDS	MI			-					-	-		+	+		+	-						-		+	+	+
GRO	GREENSBORO	NC											+		-	+	+		-			-		-	-		+
GRE	GREENVILLE	SC	-		+			+	-				+			+	++		-	-		--		-			+
HRG	HARRISBURG	PA		-						-		-		+	+			-					-	-	-	-	
HRD	HARTFORD	CT								-	-		-	+	+	+				+	+			+			+
HNU	HONOLULU	HA		+	++		-		--		+++	+	-	-		-		+++	+		+		+++	+			-
HSN	HOUSTON	TX	+		++	+	+		-		+	++			-			+		+	+	-	+		-	-	
HNN	HUNTINGTON-AD	WV	-			-	++		+	+		-	+	--			++		-	--	-		-	-	+		
INS	INDIANAPOLIS	IN	+							-	-			+								-				+	
JCN	JACKSON	MS	-	+			-	+		++			+	--	-	-	+	+	+	-	-			-		-	
JCE	JACKSONVILLE	FL			+		-			+		++			-	-				-	-		-		-	-	-
JC	JERSEY CITY	NJ			-	-	+	-	+			-	---	+	+	+	+	+	--	-	-	+			+		
JHN	JOHNSTOWN	PA	-	-	-	-			+		-	--		+	+		++		-	-	--		--	--	+		
KC	KANSAS CITY	MO	+									+		-							+						
KNE	KNOXVILLE	TN		-						+			+	-			++			-			-	-			
LNK	LANCASTER	PA	-				+			-		-		+	+	+	+	-	-		-	-		-	-	-	+
LNG	LANSING	MI		+	-		+	+	-	-	-	-		+					+	+	+			+	+	++	+
LV	LAS VEGAS	NV	-	+		+++		-	--		+		-	-	--	--	-		+				+	++	+		-

Table A-11 (Cont.)

DATA SUMMARY

			POPUL	BIRTH	CONEM	POPCH	CAPEX	VARAT	OLD	GINI	CLIM	DENSC	1 UNIT	POP 20	AGECI	MANUF	-PLUM	CROWD	HIGHS	INCOM	COLLG	WELF	BLFAM	RENT	UNEMP	UNINC	VAINC
LR	LITTLE ROCK	AR	-				-			+				--			+			-				+	-		
LRN	LORAIN-ELYRIA	OH	-				+	-	-	--			+	+		+				+	-	-	+				--
LA	LOS ANGELES	CA	++		-					+	++	++	-				-		+			++		+	+		
LSE	LOUISVILLE	KY					+							-	+	+			-				-				+
MDN	MADISON	WI	-			+	-				-		-	+	-	-			++	+	++	-		+	-		-
MMS	MEMPHIS	TE		+						+		+		-			+	+	-	-		+		-			
MMI	MIAMI	FL	+	-	+	+	-	+	++	++	+		-			-		++						++			-
MLE	MILWAUKEE	WI	+		-					-	-	+	-	+		+											
MNS	MINNEAPOLIS	MN	+							-	-			-	+				+	+	+		-	+	-	+	
MBE	MOBILE	AL			+	-				+		+	+		-		+	+		-	--			-		-	
NSE	NASHVILLE	TE					+			+		++		-			+		-			-			-	-	
NH	NEW HAVEN	CT								-	-		-	+	+		-			+	+				+	+	
NO	NEW ORLEANS	LA								++		+	-		+	-		+	-	-		+			+		
NYC	NEW YORK	NY	+++	-	-	-	-	-	+	+		++	---	+	+							+					
NWK	NEWARK	NJ	+	-			+			+			-	+	+					+	+	+		+			+
NN	NEWPORT NEWS	VA	-	+			-	-	--					+	-								+				-
NRK	NORFOLK	VA		+	+		-	-	-					+	-	-			-	-					-	-	-
OC	OKLAHOMA CITY	OK						-				++	+	-		-	-		+		+					-	-
OHA	OMAHA	NE									-			-					+				-		-		
ORO	ORLANDO	FL			++	+		-		+	+	-	+		-	-									-		-
OXD	OXNARD-VNTURA	CA				+++		+++	-		++	-	+		-	-	-		+	+			+	+	+	+	
PTN	PATERSON	NJ	+	-				-			-	-	-	+	+	+	-	-		+				+			
PRA	PEORIA	IL	-							-	-	-	+	-		+						-				-	
PHA	PHILADELPHIA	PA	++									+		+	+		-	-				+					
PHX	PHOENIX	AR			+	+		+			+	+		-	-			+				-		+			
PTG	PITTSBURG	PA	+	-		-			+					+	+								-	-			
PRD	PORTLAND	OR		-					+					-				-	+						+	+	
PRE	PROVIDENCE	RI		+	-	-			+				-	+	+	+			-		-		-	-			
RDG	READING	PA	-	--	-	-	+		+	-		-		+	+	+	+	-	-		-	-		-	-		+
RCD	RICHMOND	VA			+		+							+					-						-	-	
RCR	ROCHESTER	NY			-		+	+		-	-			+	+	+		-		+				+			+++
RKD	ROCKFORD	IL	-	+	-		+	+		-	-			+		+					-					+	+++
SCO	SACRAMENTO	CA		-				-	-		+				-	-	-		+			++	+	+	+	-	-
SL	ST LOUIS	MO	+											-	+			+	-			+	-			+	
SLS	SALINAS-MNTRY	CA	-		-		+	++	-		++	-			--	-	-	+	+	-	+	+	+	+	+		
SLC	SALT LAKE CTY	UT		++				-	-					-		-	-	+	+		+		--		-	-	-
SA	SAN ANTONIO	TX		+			-			+	+	+	+		-	-	+	++	-	-							-
SB	SAN BERNARDNO	CA	+			+			+		+		+		-	-	-					+	+		+		-
SD	SAN DIEGO	CA	+			+	--	-			++	+			-	-	-		+		+			+	+	+	-
SF	SAN FRANCISCO	CA	+	-				-			+		-		+	-			+	+	+	+		+			-
SJ	SAN JOSE	CA				++		+	-	-	+				--		-		+	++	++	+	++	++	+		
SBA	SANTA BARBARA	CA	-			++	-	+			++	-			-	-	-		++	+	+		+	+			-
SEE	SEATTLE	WA	+							-	+			-				-	+	+	+			+	++	+++	
SHT	SHREVEPORT	LA	-	+	+	-	-	++		+			+			-	++	+		-				-			
SBD	SOUTH BEND	IN	-			-				-	-		+	+								-			+	+	
SPE	SPOKANE	WA	-	-		-		-	+				+	--	-	-		-	+	-				-	+		-
SPD	SPRINGFIELD	MA		-					+		-		-	+		+											
STN	STOCTON	CA	-		-						+		+					+	-		-	++	-		++		
SYE	SYRACUSE	NY						-			-			+				-					-			+	
TCA	TACOMA	WA					-							-		-	-	-	+				+		++	++	
TMA	TAMPA	FL		--	+	+	-		+++	+	+				-	-		-		-					-		-
TLO	TOLEDO	OH			-					-				+							-						
TRN	TRENTON	NJ	-		-			-				-		+	+		-	-		+	+	+		+			-
TCN	TUCSON	AZ	-		++	+	--	-		+	+			-	-	-		+	+	-	+		+		-	-	-
TLA	TULSA	OK						+					+	-	-								-				
UTA	UTICA-ROME	NY	-		-	-			+	-	-	+	-	+	+	+		-	-		-			-	+	+	+
WSN	WASHINGTON	DC	+	+		+	-		-	+			-	+		--	-		+	++	++	-		++	-	-	-
WPB	WEST PALM BCH	FL	-	-	++	+	-	+	+++	++	+	-													-		
WCA	WICHITA	KA		+		-							+	--	-		-		+	-		+			++	+++	+
WB	WILKES-BARRE+	PA	-	--		-	-		+		-	-	-	+	+	+	+	-	-	--	--		---	-			
WLN	WILMINGTON	DE			+		+	-				-		+	+	+		-									
WRR	WORCESTER	MA		-			-		+	-	-		-	+	+	+								-			
YRK	YORK	PA	-		+					-		--	+	+	+	+	+	-	-		-			-	-		+
YNN	YOUNGSTOWN	OH		-		-	+			-	-		+	+		+					-	-				+	+
TAG	-----CITY------		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25

Table A-11 (Cont.)

DATA SUMMARY

TAG	-----CITY------		PCH 50 26	DENSS 27	BOWNH 28	SEGSC 29	INCSC 30	C% POP 31	CROSC 32	RENSC 33	BLINC 34	INCBW 35	INFBW 36	BLSPA 37	SPANI 38	VACAN 39	ROB 40	BURG 41	OEO$ 42	MGSC 43	INFMO 44	COSTL 45	INCGR 46	SEGBL 47	DENTI 48	GOVT 49	#GOV 50	LOCEX 51	FEDEM 52	LBJ64 53
AKN	AKRON	OH								-	+				-				-				-			-			-	+
ALY	ALBANY	NY	-		-			-					+	-	-		-	-	-					-			+	+		+
ALE	ALBUQUERQUE	NM	+		-	-	-	+	--	++		-	-	+	++	-		+	++			-	--	-			--		+	
ALN	ALLENTOWN	PA			-	-		-				-		--	-	-	-	-			-	+		-		+	+	-	-	+
ANM	ANAHEIM	CA	+++		-	-		-	+			+	-	-	+			+	--	-	+		-			+		+	-	-
APN	APPLETON-OSKS	WI			--	-	-		-	+			+	--	-	-	-	-	-	+	-								-	
ATA	ATLANTA	GA			+	++	+		+	--		+		+					+	-	+	-	+	+						-
AGA	AUGUSTA	GA	-	-	+	+	++	-	+	-	-	+		+				-		-	+	-	+++	+			-	-	+	-
ASN	AUSTIN	TX		+		-		++		+		+				+						-	+	+		+				+
BKD	BAKERSFIELD	CA	+	-			-	-	-	+	-	++			++	+			+	+	-				-	+	+	++	+	
BLE	BALTIMORE	MD	-	+	+	++	+		+	-	+		-	+			+++			-					-	-	--		+	
BR	BATON ROUGE	LA			++	-		+	-	+	-	++				+		+	+	-		-	-	+		-	--		-	-
BMT	BEAUMONT	TX		-	+	+					-	+				+			-			-		+	-	+			-	
BNN	BINGHAMTON	NY	-	-	-	-		-			++	--	++	-		+	-	-	-		--					-	+	+	-	
BRM	BIRMINGHAM	AL			++	+				+	-			+	-		-					-		+		--		-		--
BSN	BOSTON	MS	-		-		+	-	+		-	+										++			+					+
BRT	BRIDGEPORT	CT			-		+	-	+		+		-			-					-	+	+	-	+				-	
BFO	BUFFALO	NY	-	+			+						+					-				+	-		+			+		+
CNN	CANTON	OH	-								+	-	++	-				-	--		+				-			-	-	
CHN	CHARLESTON	SC	-		++		-	-		-	-	+		+		+			+		+	-	++	-	-			-	++	--
CHE	CHARLOTTE	NC	+		+	+		+		+	-	+	+		-			+			+		+	+	-	+	-			
CHA	CHATTANOOGA	TN		-		+	+		+		-	+	-	+	-				+		+	-	+	+	-		-			-
CHO	CHICAGO	IL	-	++		+	+		+	-	+			+			+	-		-	+	+		+	+		+			
CNI	CINCINNATI	OH				+							-		-	+		-		-					-	+	+			
CLD	CLEVELAND	OH	-	+	+	+	+		+	--	+			+			+	-		-		+		+	+	-				+
CLA	COLUMBIA	SC			+						-					+					++	-	+	+	-	+		-		-
CLS	COLUMBUS	OH		+				+			+	-	-		-								-		+					
CC	CORPUS CHRSTI	TX	+	+		-	-	+	--	+	-		---	+	++	+		+		++		--	+		-	+			+	+
DLS	DALLAS	TX	+		+	+						+				+		+				-		+		+				
DVT	DAVENPORT-RI+	IA		-	-	-	-			+			-		+			-	-		-		-		-		+		+	
DYN	DAYTON	OH				+	+		+		++	-				-	+			-				+	-				+	
DNR	DENVER	CO								-					+			+	+	-			-		+	-	+		+	
DM	DES MOINES	IA				-		+		-								-	+	-		+	-		+	+				
DTT	DETROIT	MI	-	+	+	+	+			--	+	-		+			+++	+					+					+		+
DLH	DULUTH	MN		--	-	-	-		-	+		-		--	-		-	-	++		-				+		+	+		+
EP	EL PASO	TX	+		-	-	-	++	-	+	+	--	-	++	++					+		--		-	-	-	-	-	+	
ERE	ERIE	PA			-									-	-			-			-					-		-	-	+
FLT	FLINT	MI			+	+					+	-	+						-	-		+		+	-	+			-	
FL	FORT LAUDRDLE	FL	++			-	-	-					++			++		+		+	+	-		++	+	+			-	-
FWE	FORT WAYNE	IN						+		-	+	-	+	-	-					-		+		+	-	--		-	-	-
FW	FORT WORTH	TX		-						-									-	-	+	-	-	+	-	+		-		
FRO	FRESNO	CA				-	-		-	+	-	+			++			+	+	+	-					+	+	++		
GRY	GARY	IN			+	++			+		+	-		++			+				+			+	-				-	
GR	GRAND RAPIDS	MI							-					-					-							+			-	
GRO	GREENSBORO	NC			+	+	-	-		+			+		-					+	+		+	+	-	+	-		-	-
GRE	GREENVILLE	SC		--				-		+	-		+		-					-	++	-	++		-	+		--	-	-
HRG	HARRISBURG	PA	-				+	-					+			+		-	-							-	+		+	
HRD	HARTFORD	CT			-		++	-	+	-			+			-						+			+	+				+
HNU	HONOLULU	HA			-	++	-			+	-	+	--	-				+			--	++	++	-	+		---	---	++	+
HSN	HOUSTON	TX	+		+			+		+		+		+		+	+	+				-	-	+		-				
HNN	HUNTINGTON-AD	WV	-		-	-	-		-	+		-		-	-		-	-				-			-	+		-		
INS	INDIANAPOLIS	IN	+	++			-	+			+	-			-				+		+	+		+			+			-
JCN	JACKSON	MS			++		-	+		+	-	++		+	-	-	-	-	++		+	-		+	-		-			---
JCE	JACKSONVILLE	FL		--	+	-	-	+++	+++		-	+			-	+	+	++		+		-		+	-		--		+	-
JC	JERSEY CITY	NJ	-	++								-			+			-	+		+	+		-		-	-			+
JHN	JOHNSTOWN	PA	-	-	-	-		-		+	-		+	-		---	-	--			-			-	-		+	-	-	
KC	KANSAS CITY	MO			+								+			+	+		+						+	+	+			
KNE	KNOXVILLE	TN	-							+	-	+	-	-	-		-	-		+++		-				--	-	-		-
LNR	LANCASTER	PA	-		-			-	+					-			-	-	-		-			-	-	-	+	-	-	-
LNG	LANSING	MI									+	-		-				+									+		-	
LV	LAS VEGAS	NV	++			-	-		-								+	+			-		--	+	-	+	-	+		
LR	LITTLE ROCK	AR			+		-		-	+	-	+			-				+			-				+		-		

Table A-11 (Cont.)

DATA SUMMARY

TAG	CITY	ST	PCH 50	DENSS	BOWNH	SEGSC	INCSC	C% POP	CROSC	RENSC	BLINC	INCBW	INFBW	BLSPA	SPANI	VACAN	ROB	BURG	OEOS	MGSC	INFMO	COSTL	INCGR	SEGBL	DENTI	GOVT	#GOV	LOCEX	FEDEM	LBJ64
LRN	LORAIN-ELYRIA	OH									+	-	+					-	-					-	-				-	+
LA	LOS ANGELES	CA		+									-	+	+		+	+				+				--		+		
LSE	LOUISVILLE	KY				+				-					-					-										
MDN	MADISON	WI			-	+	-	+					++	+	-	-	-	-	+	+	--		-		+					+
MMS	MEMPHIS	TE		+	++		-	++	-	+	-	+		+	-	-				+++	+	-	+	+	+		-			
MMI	MIAMI	FL					+	-	++	-				++	++	-	++	+			+	-	+	++		+	-			
MLE	MILWAUKEE	WI		+												-	-	-				+	-		+			+		
MNS	MINNEAPOLIS	MN	-		-	-	+	-	-	--				-		-			-	-		+		-	+			+		
MBE	MOBILE	AL	+	--	++		-		-	+	-	+		+	-			+			+	-		+	-		-	-	-	--
NSE	NASHVILLE	TE		--	+		-	++		++			-		-				+		+	-	+	+			-			
NH	NEW HAVEN	CT						-	+								-					+		-	+					+
NO	NEW ORLEANS	LA			+	+	+	+		-	-	+	+	+		+	++			-	+	-				-	-			-
NYC	NEW YORK	NY	-	+++				+	+	-						-	+++	+				++		-	+++			++		+
NWK	NEWARK	NJ	-	+		+	+++	-	++	-			+	++	+		+				+	+		-	+	-		+		
NN	NEWPORT NEWS	VA			+								+		-						+	-			-	+	--	-	++	
NRK	NORFOLK	VA		+	+					-										-	+	-	+	+	-	+	--		++	
OC	OKLAHOMA CITY	OK		--				+								+						-				+			++	
OHA	OMAHA	NE		+			-	+						-										+			+			
ORO	ORLANDO	FL	+	-	+			-								+						-		++		-				-
OXD	OXNARD-VNTURA	CA	+		-	-				+	+				++		-				--	-	---			--		+	+	
PTN	PATERSON	NJ		+			+++	-	+	-	+			+		--		-	-		-	+		-	+				-	
PRA	PEORIA	IL	-				-						+			+			-	+					-	+	+	-	-	
PHA	PHILADELPHIA	PA	-	+	+	+				-	+							-			+					-	+			
PHX	PHOENIX	AR	+++			-	-	+			-	+	-		+			+	+				+			+				-
PTG	PITTSBURG	PA	-					-		+								-								-	+			+
PRD	PORTLAND	OR			-	-			-					-				+			-	+		-	++		+			
PRE	PROVIDENCE	RI	-		-		+	-			-	+	+	-	-						-			-				-		++
RDG	READING	PA	-		-									-			-	-	-					-			+	-	-	
RCD	RICHMOND	VA	-		+	+			+	-			-	+	-			+	-	+	+		+	+	+	+	--			-
RCK	ROCHESTER	NY	-				+			-			+				-	-		-	-	+			+	+	+	+	-	+
RKD	ROCKFORD	IL							-				+		+			-	--	+		+			-		+	-	-	
SCO	SACRAMENTO	CA								-					+			+		+	-		-	--		+	+	+	+	
SL	ST LOUIS	MO	-			+	+	-	+	-				+		++	+			-	+					-	+			+
SLS	SALINAS-MNTRY	CA	+				-		-	+		-			++	-			+	++	-				+	+		+		
SLC	SALT LAKE CTY	UT		-	-	-			-				-	-					++	-	-		-	--					+	
SA	SAN ANTONIO	TX				-		+	+	--				+	++				+	-		--	+		-	+		-	++	
SB	SAN BERNARDNO	CA	+	-		-		-					-		++	-		+				-	-				+	++		
SD	SAN DIEGO	CA	+			-	-					-	-		+							+		-		+		+	+	-
SF	SAN FRANCISCO	CA	-	+		-		-					-		+		+	++				+		--	++			+	+	
SJ	SAN JOSE	CA	+	+	-	-					+++	--	--		+	-				++	-			--	+	+		+		
SBA	SANTA BARBARA	CA		-	-	-		-	-	+			-		++	+	-		-	-	--		--		+	+		+		
SEE	SEATTLE	WA			-						+	-	-	-		+		+		-	-	+		-	++		+			
SHT	SHREVEPORT	LA			++	-	-	+	-		--	++	-			+	-	-		+	+	-		+			-			---
SBD	SOUTH BEND	IN					-			+		-	+			+					-		--						-	
SPE	SPOKANE	WA			-	-		+	-			-	-	-	-				-		-	+		-	+	+				
SPD	SPRINGFIELD	MA			-								++				-						-							+
STN	STOCTON	CA				-	-			+	-		-		+			+	+	+				-		+	+	++	+	
SYE	SYRACUSE	NY	-		-								-	-				-						-			+	+		
TCA	TACOMA	WA		-		-				-	+	-		-		+	-							-		+				
TMA	TAMPA	FL	+					-		-	-							+			+	-		+		+	-			-
TLO	TOLEDO	OH		+				+	-		+					-				+				+	-	+			-	+
TRN	TRENTON	NJ	-	+	+	+	+		+	-	+		-	+		-	+	+	+	-			-	-		-				+
TCN	TUCSON	AZ	+++		-			+	-	+			+		+				+	-		-		-		+	-			
TLA	TULSA	OK		+		-	-	+	-	++	-	++		-	-	+			+	++	+							-		-
UTA	UTICA-ROME	NY		--	-					+		-	+	-	-		-	-	-					-		-	+	+		
WSN	WASHINGTON	DC	-	+	+	++	++	-	++	-	+			++			++		+	-		+		-		-	-	+	+++	+
WPB	WEST PALM BCH	FL		--				-			-	+	+		+	+		+		-	+	-		++	+	+		+	-	-
WCA	WICHITA	KA	+	+			-	+	-	+				-		+				+			-	+	-	+	+			
WB	WILKES-BARRE+	PA	-		-	-		-		+	--	+		--	-		-	--	+		+		+	-			+	--		+
WLN	WILMINGTON	DE	-	-		+	++	-		-			+	+		+				-	-	+	-	-	-					
WRK	WORCESTER	MA	-	-	-	-				+				-	-	-			-							+			-	+
YRK	YORK	PA	-		-			-						-				-	-		---			-			+	-	-	
YNN	YOUNGSTOWN	OH				+		-			+	-				-		-	-					-	-			-	-	+
TAG	-----CITY------		26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53

Table A-12

CORRELATION MATRIX

% CORR	POPUL	BIRTH	CONEM	POPCH	CAPEX	VARAT	OLD	GINI	CLIM	DENSC	1 UNIT	POP 20	AGECI	MANUF	-PLUM	CROWD	HIGHS	INCOM	COLLG	WELF	BLFAM	RENT	UNEMP	UNINC	VAINC	PCH 50	DENSS	BOWNH	SEGSC	INCSC	C% POP
1	100																														
2	-14	100																													
3	-15	7	100																												
4	4	16	37	100																											
5	5	-10	-36	-24	100																										
6	-13	5	26	36	-14	100																									
7	7	-72	-2	-22	-2	12	100																								
8	2	11	54	16	-50	21	6	100																							
9	2	18	42	59	-35	29	-20	50	100																						
10	54	19	6	2	-18	-9	-16	22	14	100																					
11	-50	16	23	0	7	12	-22	6	16	-10	100																				
12	27	-29	-35	-4	29	-8	17	-39	-18	-21	-32	100																			
13	21	-23	-41	-62	31	-31	27	-34	-68	-20	-34	22	100																		
14	-1	-19	-45	-42	66	-3	20	-60	-59	-25	-3	38	49	100																	
15	-36	8	17	-46	-6	0	5	29	-22	-14	22	-38	22	9	100																
16	-11	54	36	9	-28	13	-38	56	49	24	13	-40	-42	-41	28	100															
17	9	12	3	46	-25	4	-20	-8	28	14	-3	-15	-33	-46	-51	-12	100														
18	45	4	-14	42	22	6	-25	-43	-4	12	-28	37	1	10	-62	-26	39	100													
19	27	17	16	45	-26	4	-31	11	28	21	-21	-8	-22	-47	-39	-5	66	51	100												
20	19	2	-18	-7	4	-15	1	23	27	4	-14	8	-2	-22	-25	11	9	-8	3	100											
21	-6	35	22	46	-1	22	-39	-1	42	8	8	0	-44	-11	-21	32	22	33	19	-11	100										
22	41	3	4	70	-11	19	-16	-3	34	13	-36	18	-28	-30	-66	-5	54	73	61	7	37	100									
23	-1	9	-35	-6	15	-1	-9	-13	15	-9	-1	2	-3	4	-22	8	12	0	-6	39	5	3	100								
24	9	5	-35	1	17	8	-4	-37	-11	-1	0	8	8	29	-21	-17	12	23	-1	-4	7	18	62	100							
25	-2	-8	-25	-28	48	38	12	-38	-37	-11	3	18	32	67	3	-26	-26	15	-22	-18	-3	-13	5	23	100						
26	-11	24	39	64	-33	26	-19	26	61	22	19	-21	-82	-41	-24	35	33	2	20	-4	42	31	2	-6	-27	100					
27	45	4	-19	-4	3	-8	-5	-7	-11	0	-43	19	14	5	-22	-4	2	29	17	11	-3	31	5	13	0	-7	100				
28	8	24	32	-2	1	-1	-29	47	4	26	24	-12	-6	-12	26	33	-30	-5	-9	4	3	-11	-19	-15	-7	-1	1	100			
29	35	3	-2	-15	26	-20	-21	-1	-18	14	-19	21	30	14	-3	9	-23	35	9	-1	8	13	-18	-1	12	-29	21	45	100		
30	42	-18	-19	-6	23	-14	7	-14	-27	-13	-38	37	51	24	-20	-30	-9	45	17	7	-13	27	-8	2	14	-42	18	7	46	100	
31	-11	44	13	-3	-30	0	-32	26	11	59	15	-41	-37	-31	12	39	8	-17	9	-2	9	-8	-11	-9	-16	31	16	24	-13	-45	100
32	44	-13	-8	7	9	-5	0	0	-8	19	-39	37	21	12	-16	-13	-16	39	11	1	2	32	-21	-9	7	-14	15	25	41	61	-6
33	-52	5	21	-3	-11	14	-3	12	17	-12	35	-24	-39	-8	29	21	-1	-42	-15	-8	3	-42	2	-15	-8	30	-26	-15	-40	-63	25
34	27	0	-35	17	29	-9	-15	-60	-15	0	-19	38	4	28	-46	-34	27	69	23	-1	40	50	13	36	13	-4	24	-26	14	28	-15
35	-5	7	37	9	-22	21	-3	52	20	6	8	-24	-9	-31	16	30	-5	-19	7	1	-27	-13	-15	-30	-6	10	-12	35	6	-8	8
36	-13	-22	-2	-10	9	-2	27	-15	-44	-23	-13	28	23	29	-1	-29	-19	0	-19	-14	-8	-11	-13	2	14	-16	-8	2	7	14	-20
37	29	24	20	11	-8	0	-25	45	22	17	-7	5	-9	-21	3	43	-19	20	17	15	14	24	-10	-21	-13	7	18	63	56	39	6
38	4	17	6	37	-20	19	-10	25	66	-1	-1	-3	-48	-38	-31	42	20	7	21	32	30	38	26	-14	-22	45	7	-22	-26	-9	4
39	-1	18	45	14	-16	19	-4	28	22	15	16	-23	-9	-23	0	17	1	-6	9	5	9	0	-5	-1	-5	9	-11	31	2	-14	15
40	66	0	0	6	1	-12	-6	22	6	42	-32	9	6	-17	-25	13	-5	34	16	20	-1	37	-3	0	-14	-8	41	36	45	33	12
41	24	19	31	46	-24	12	-21	44	59	38	3	-21	-45	-53	-30	35	28	16	39	26	22	41	16	2	-31	40	1	26	0	-9	23
42	-3	16	9	-10	-33	-13	-9	50	19	17	-1	-36	-14	-47	23	40	5	-33	13	24	-29	-10	0	-28	-40	9	4	21	5	-21	28
43	-22	-2	-1	-7	-7	18	0	10	12	-9	19	-11	-32	-7	10	14	-4	-25	-9	10	0	-16	15	-6	-3	15	11	-9	-31	-48	21
44	11	7	21	-4	-7	6	-1	38	-2	24	1	-16	-5	0	24	25	-37	-14	-17	-11	-8	-13	-14	-10	-1	9	7	61	25	2	17
45	36	-26	-50	-12	27	-23	11	-61	-29	-6	-49	36	40	27	-47	-48	32	43	18	10	0	30	22	30	20	-30	34	-53	8	30	-28
46	0	0	16	-24	-6	1	0	30	-6	3	-1	-15	22	12	49	20	-43	-28	-19	-12	-9	-24	-18	-4	11	-19	-11	33	22	10	-2
47	-10	10	42	12	-8	23	1	37	0	21	25	-22	-25	-11	25	27	-23	-15	-15	-19	0	-7	-35	-7	1	16	-12	57	18	-17	29
48	49	-38	-17	6	-5	0	33	-4	0	11	-54	13	18	-12	-34	-32	43	30	45	21	-11	42	8	11	-4	-17	34	-33	-1	29	-17
49	-25	-3	14	21	-20	20	0	17	23	-2	23	-22	-36	-18	-4	9	7	-16	1	-11	7	5	6	0	0	32	-22	-2	-25	-32	3
50	13	-32	-45	-22	32	1	33	-40	-31	-20	3	17	31	32	-9	-45	6	8	-14	5	-27	-9	14	20	27	-28	0	-41	-20	5	-32
51	28	-10	-25	25	0	-3	10	0	24	11	-23	24	-24	-26	-50	-10	31	27	26	52	2	41	30	6	-13	21	9	-21	-15	6	-9
52	10	34	11	16	-45	-12	-42	31	32	22	-1	-23	-10	-62	-3	26	23	0	35	14	7	17	-5	-25	-44	5	0	18	12	3	18
53	24	-19	-40	-17	13	-22	9	-46	-22	1	-42	27	30	17	-36	-26	2	29	11	5	-4	20	12	15	11	-31	22	-53	6	23	-12

Table A-12 (Cont.)

CORRELATION MATRIX

% CORR	CROSC	RENSC	BLINC	INCBW	INFBW	BLSPA	SPANI	VACAN	ROB	BURG	OEO$	MIGSC	INFMO	COSTL	INCGR	SEGBL	DENTI	GOVT	#GOV	LOCEX	FEDEM	LBJ64
32	100																					
33	-52	100																				
34	21	-30	100																			
35	-1	13	-81	100																		
36	8	-8	7	-10	100																	
37	44	-30	0	17	-7	100																
38	-6	7	7	-3	-31	33	100															
39	-9	9	-21	29	0	12	1	100														
40	44	-46	19	-1	-11	51	10	7	100													
41	12	-7	-5	21	-36	35	39	21	50	100												
42	-10	15	-37	29	-21	26	13	4	14	28	100											
43	-32	50	-13	3	-20	-13	18	-3	-21	1	11	100										
44	17	-12	-26	25	8	43	-10	17	26	14	15	-7	100									
45	13	-29	45	-32	3	-37	-10	-28	8	-17	-27	-9	-43	100								
46	17	-14	-30	21	5	14	-30	-2	8	0	9	-5	38	-19	100							
47	7	5	-28	27	17	22	-20	33	10	11	1	-6	46	-51	20	100						
48	13	-30	20	-7	-12	-13	4	-17	21	10	1	-5	-26	59	-10	-37	100					
49	-22	9	-21	16	-8	-9	15	3	-17	19	5	14	18	-18	14	11	-13	100				
50	-16	-1	19	-22	10	-32	1	-6	-13	-36	-19	4	-31	34	-27	-26	22	-6	100			
51	3	-9	29	-17	-12	10	46	-8	25	31	10	11	-19	22	-38	-21	33	5	23	100		
52	8	-13	-8	10	-27	29	23	16	21	26	36	-6	6	-15	6	-10	-4	2	-29	4	100	
53	8	-19	41	-39	5	-20	10	-21	14	-10	-12	-3	-36	55	-16	-43	27	-9	25	23	-1	100

Table A-13

FACTOR SCORES

CITIES		Non Manu-facturing	Inequality	Ghetto-Suburbs	Growth	Stage in Life Cycle	Welfare	Density	Education
AKRON	OH	-0.9	-0.7	0.4	-0.0	0.2	-0.8	-0.1	-0.1
ALBANY	NY	0.7	-1.0	0.0	-0.7	-1.1	-0.8	-0.1	0.4
ALBUQUERQUE	NM	2.1	-0.9	-1.5	0.6	1.4	0.5	0.3	0.3
ALLENTOWN	PA	0.0	-1.5	-1.2	-0.7	-1.4	-1.2	-0.2	-1.1
ANAHEIM	CA	-1.4	0.3	-0.1	3.2	-0.6	0.0	-0.2	1.6
APPLETON-OSKS	WI	-1.0	-1.1	-1.8	0.2	-0.2	-1.5	0.1	0.1
ATLANTA	GA	0.4	0.8	2.3	0.3	0.6	-0.4	-0.8	0.6
AUGUSTA	GA	-0.2	1.5	1.6	-1.3	0.7	0.7	-2.0	-0.8
AUSTIN	TX	1.5	0.7	-0.8	0.9	0.6	-1.5	1.0	0.7
BAKERSFIELD	CA	1.1	0.2	-0.8	0.3	-0.6	3.1	-1.6	-0.8
BALTIMORE	MD	0.4	-0.1	1.9	-0.3	0.7	0.1	1.2	-0.8
BATON ROUGE	LA	-0.1	1.7	-0.2	0.1	1.3	0.2	-0.2	0.7
BEAUMONT	TX	-1.6	1.6	0.3	0.3	0.1	0.2	-1.5	-0.9
BINGHAMTON	NY	-0.8	-2.1	-0.3	0.1	-0.6	0.1	-0.9	-0.6
BIRMINGHAM	AL	0.4	0.9	0.2	-0.8	-0.0	-0.3	-0.2	-1.6
BOSTON	MS	0.8	-0.1	0.6	-1.2	-1.3	0.3	0.8	1.6
BRIDGEPORT	CT	-1.2	-0.6	0.5	0.1	-0.5	-0.0	0.3	0.5
BUFFALO	NY	-0.5	-0.5	0.3	-1.0	-0.8	0.6	0.9	-0.2
CANTON	OH	-1.2	-1.0	-0.1	0.1	-0.1	-1.1	-0.2	-0.9
CHARLESTON	SC	0.9	1.7	0.4	-1.7	1.5	0.3	-0.4	-0.3
CHARLOTTE	NC	0.3	1.0	-0.1	0.3	0.7	-1.1	0.1	-0.3
CHATTANOOGA	TN	-0.4	1.5	0.5	-1.1	-0.1	0.3	-0.8	-0.6
CHICAGO	IL	-0.4	-0.0	1.2	-0.2	0.1	-0.2	2.3	0.1
CINCINNATI	OH	-0.1	-0.2	0.6	-0.6	0.0	-0.4	-0.1	-0.6
CLEVELAND	OH	-0.4	-0.2	1.8	-0.3	-0.2	-0.1	0.8	0.0
COLUMBIA	SC	0.3	1.5	-0.1	-0.3	0.7	-0.9	-0.3	-0.2
COLUMBUS	OH	0.5	-0.8	0.2	-0.2	0.7	-0.8	0.7	0.4
CORPUS CHRSTI	TX	0.7	0.8	-2.2	0.6	1.5	0.8	1.2	-1.3
DALLAS	TX	-0.4	1.2	0.4	1.1	0.4	-0.4	0.0	0.7
DAVENPORT-RI+	IA	-0.7	-0.5	-0.8	0.1	0.1	-0.3	-0.7	-0.1
DAYTON	OH	-0.9	-0.6	1.3	0.3	0.8	-0.7	-0.2	-0.3
DENVER	CO	1.0	-0.6	0.4	0.1	0.1	-0.1	0.1	1.2
DES MOINES	IA	1.0	-0.5	-0.3	-0.5	0.1	-1.0	-0.2	1.1
DETROIT	MI	-1.8	0.4	2.0	-0.1	1.1	1.3	1.3	0.1
DULUTH	MN	1.4	-1.4	-1.7	-1.8	-0.4	-0.2	-1.3	0.8
EL PASO	TX	1.2	-1.0	-1.3	1.0	3.3	0.1	0.4	-2.1
ERIE	PA	-0.5	-0.7	-0.8	-0.8	-0.3	-0.8	0.2	-0.2
FLINT	MI	-3.0	-0.1	0.5	0.3	1.7	1.2	-0.2	-0.6
FORT LAUDRDLE	FL	0.0	2.4	-0.8	3.4	-3.3	-2.1	1.4	0.4
FORT WAYNE	IN	-1.2	-0.6	-0.2	0.2	1.0	-1.3	0.1	0.3
FORT WORTH	TX	-1.6	1.1	0.6	1.0	0.5	-0.5	-0.6	0.1
FRESNO	CA	0.9	0.0	-1.4	0.1	-0.7	3.2	-0.0	-0.8
GARY	IN	-0.8	-1.0	1.2	0.5	1.5	-0.1	0.0	-2.4
GRAND RAPIDS	MI	-1.5	0.1	-0.4	-0.3	-0.2	0.2	0.0	0.4
GREENSBORO	NC	-0.3	0.6	-0.5	0.2	0.2	-1.1	-0.1	-1.6
GREENVILLE	SC	-1.0	1.2	0.2	-0.1	0.2	-1.1	-1.8	-1.3
HARRISBURG	PA	0.7	-0.5	0.7	-0.7	-1.1	-0.8	-0.9	-0.2
HARTFORD	CT	-0.4	-0.3	1.3	-0.2	-0.6	-0.4	-0.2	0.9
HONOLULU	HA	1.7	-0.2	-0.3	0.1	2.2	-1.1	-0.2	2.3
HOUSTON	TX	0.1	0.8	0.3	1.3	0.8	-0.8	0.3	-0.2
HUNTINGTON-AD	WV	-0.3	0.1	-1.5	-1.3	0.0	0.3	-0.2	-0.8
INDIANAPOLIS	IN	-0.7	-0.0	-0.7	-0.0	0.9	-0.8	2.1	0.1
JACKSON	MS	0.8	2.2	-0.9	-0.7	0.8	-0.2	-0.4	-0.1
JACKSONVILLE	FL	1.7	1.1	0.2	0.1	0.6	-0.6	-1.0	-0.5
JERSEY CITY	NJ	0.6	-1.6	-0.3	-0.8	-0.2	0.7	3.0	-2.6
JOHNSTOWN	PA	0.2	-1.1	-0.9	-1.6	-1.4	0.3	-0.9	-1.7
KANSAS CITY	MO	-0.3	0.5	0.6	-0.5	0.1	0.3	-0.3	0.9
KNOXVILLE	TN	0.4	0.7	-1.5	-1.2	-0.1	-0.4	0.7	-0.4
LANCASTER	PA	-0.2	-1.2	-0.0	-0.2	-1.2	-1.5	-0.7	-1.2
LANSING	MI	-1.6	-0.4	-0.2	0.3	0.9	0.4	-0.1	1.1
LAS VEGAS	NV	0.5	-0.2	-0.3	2.0	1.0	0.2	-0.6	1.2

TABLE A-13 (cont.)

FACTOR SCORES

CITIES		Non Manu-facturing	Inequality	Ghetto-Suburbs	Growth	Stage in Life Cycle	Welfare	Density	Education
LITTLE ROCK	AR	0.5	1.9	-0.9	-0.9	0.3	-0.2	0.1	0.5
LORAIN-ELYRIA	OH	-0.5	-2.0	-0.0	0.6	1.2	-1.0	-0.6	-1.4
LOS ANGELES	CA	-0.0	-0.2	0.4	0.7	-0.0	2.3	1.1	-0.2
LOUISVILLE	KY	-0.5	0.5	0.5	-0.8	0.3	-0.1	0.2	-0.1
MADISON	WI	1.1	-1.1	-0.4	0.3	0.4	-1.9	0.7	1.5
MEMPHIS	TE	0.5	1.4	-0.9	-0.6	0.7	0.1	1.5	-0.8
MIAMI	FL	0.9	1.3	1.5	1.7	-1.8	0.8	1.1	-1.2
MILWAUKEE	WI	-0.6	-0.7	0.0	-0.5	-0.2	-0.5	1.0	0.3
MINNEAPOLIS	MN	-0.5	-0.3	0.0	-0.6	-0.1	-0.5	0.6	2.3
MOBILE	AL	0.2	1.6	-0.3	-0.5	0.6	0.5	-1.9	-1.2
NASHVILLE	TE	0.6	0.7	-0.4	-0.3	0.2	-0.7	-0.9	-0.2
NEW HAVEN	CT	-0.7	-0.6	0.5	-0.6	-0.2	0.3	-0.2	0.9
NEW ORLEANS	LA	0.9	1.5	1.3	-0.7	0.5	1.2	-0.3	-0.3
NEW YORK	NY	1.3	-0.1	0.5	-0.8	-0.8	1.2	4.4	-0.3
NEWARK	NJ	0.3	-0.2	2.4	-0.2	-1.0	1.0	1.3	-0.2
NEWPORT NEWS	VA	0.3	-0.2	0.4	0.6	1.7	-1.2	-1.1	-0.6
NORFOLK	VA	1.7	0.1	0.8	-0.0	0.8	-0.9	0.1	-0.8
OKLAHOMA CITY	OK	1.3	0.0	0.4	-0.4	0.4	-0.4	-2.1	1.0
OMAHA	NE	0.2	0.2	-1.0	-0.7	0.4	-0.7	1.1	1.2
ORLANDO	FL	1.0	0.6	0.6	0.9	-0.7	-0.9	-1.2	-0.3
OXNARD-VNTURA	CA	-0.3	-1.3	-0.5	2.9	0.5	0.6	-0.2	0.1
PATERSON	NJ	-0.1	-1.0	1.6	0.2	-1.3	-0.1	0.7	0.2
PEORIA	IL	-0.5	-0.5	-0.7	-0.2	-0.1	-0.6	0.4	-0.4
PHILADELPHIA	PA	-0.4	-0.1	1.0	-0.5	-0.3	0.2	1.4	-0.3
PHOENIX	AR	-0.1	1.1	-0.9	1.6	-0.4	0.2	-0.0	0.6
PITTSBURG	PA	-0.1	-0.2	0.1	-0.9	-1.3	-0.1	0.3	-0.2
PORTLAND	OR	-0.3	0.1	-0.5	-0.7	-0.8	0.9	-0.0	2.2
PROVIDENCE	RI	-0.1	-0.5	-0.0	-1.3	-0.6	0.1	-0.0	0.0
READING	PA	-0.7	-0.9	-0.4	-0.5	-1.4	-1.2	0.0	-0.8
RICHMOND	VA	0.4	0.7	1.3	-0.1	0.3	-1.0	0.5	-0.5
ROCHESTER	NY	-1.7	-0.0	0.7	-0.0	-0.8	-0.1	0.1	1.2
ROCKFORD	IL	-2.3	-0.1	-0.7	0.7	0.3	-0.5	0.5	-0.6
SACRAMENTO	CA	1.3	-1.4	0.3	0.7	-0.1	1.7	-1.1	0.5
ST LOUIS	MO	-0.6	0.8	1.4	-1.0	-0.0	0.9	0.4	0.1
SALINAS-MNTRY	CA	0.5	-0.8	-1.2	1.5	0.1	1.9	-0.1	-0.5
SALT LAKE CTY	UT	1.8	-1.1	-0.2	-1.1	0.7	0.4	-1.6	2.1
SAN ANTONIO	TX	1.6	0.2	0.0	0.4	1.3	0.4	0.4	-1.5
SAN BERNARDNO	CA	0.5	-0.6	0.1	1.7	-1.1	2.2	-1.4	-0.9
SAN DIEGO	CA	0.7	-0.7	-0.5	1.2	-0.3	1.1	0.4	-0.0
SAN FRANCISCO	CA	1.0	-0.4	0.7	-0.1	-0.5	1.9	1.0	1.4
SAN JOSE	CA	-0.7	-2.1	-0.2	3.1	0.9	0.7	0.6	0.2
SANTA BARBARA	CA	0.4	-0.8	-0.1	2.1	-0.9	0.8	-1.4	1.2
SEATTLE	WA	-1.7	0.1	-0.3	-0.3	0.7	1.5	0.0	2.9
SHREVEPORT	LA	-0.4	2.8	-1.4	-0.7	0.6	0.4	0.6	-0.1
SOUTH BEND	IN	-1.0	-0.6	-0.5	-0.3	0.2	-0.1	-0.1	-0.1
SPOKANE	WA	0.5	-0.6	-1.6	-0.7	-0.2	0.2	-0.1	1.4
SPRINGFIELD	MA	-0.2	-0.7	0.1	-0.4	-0.8	-0.3	-0.2	-0.2
STOCTON	CA	0.4	-0.1	-0.6	0.1	-0.5	3.2	-0.4	-0.8
SYRACUSE	NY	-0.2	-0.5	-0.2	-0.8	-0.5	0.3	-0.1	0.8
TACOMA	WA	-0.6	-0.4	-0.3	0.0	0.4	1.0	-1.5	1.2
TAMPA	FL	1.0	1.5	0.0	1.3	-2.5	-0.5	0.1	-0.9
TOLEDO	OH	-0.9	-0.4	-0.3	-0.2	0.3	-0.5	0.7	-0.3
TRENTON	NJ	0.5	-0.7	1.6	-0.5	-0.1	0.4	0.7	-0.2
TUCSON	AZ	2.2	-0.4	-1.0	1.6	-0.1	-1.0	0.1	-0.4
TULSA	OK	-0.4	1.5	-2.1	-0.4	-0.3	-0.0	1.4	0.9
UTICA-ROME	NY	-0.5	-1.0	-0.3	-0.8	-0.8	-0.0	-1.7	-0.1
WASHINGTON	DC	2.5	-0.7	3.3	0.3	1.1	-0.7	0.6	1.4
WEST PALM BCH	FL	0.5	2.4	0.6	1.6	-3.1	-0.3	-1.1	0.1
WICHITA	KA	-2.1	1.0	-1.7	-0.6	1.0	1.3	0.5	1.2
WILKES-BARRE+	PA	0.1	0.4	-1.6	-2.1	-2.3	0.1	0.1	0.0
WILMINGTON	DE	0.0	-0.4	1.9	-0.7	-0.2	0.1	-1.8	-0.1
WORCESTER	MA	-0.4	-0.8	-0.4	-0.6	-1.1	-0.4	-0.7	0.5
YORK	PA	-0.7	-0.6	-0.2	-0.4	-1.3	-1.1	-0.7	-0.8
YOUNGSTOWN	OH	-1.2	-0.6	0.4	-0.3	0.2	-0.3	-0.6	-0.9

Appendix B

METHODOLOGY

EXPLORATORY DATA ANALYSIS

The techniques we have used to classify and investigate relationships among the various metropolitan areas arise out of a branch of statistics known as "exploratory data analysis," which has the drawbacks and advantages of reliance on both statistical methodology and expert judgment. Statistical methodology enables us to reduce masses of data to a humanly comprehensible picture and to clear away known structure so that subtle relationships show up clearly. Expert judgment ensures that the analysis is confined to meaningful input and helps separate genuine relationships from statistical coincidences. Exploratory data analysis must be an interactive process to be successful.

Compared with classical data analysis, the exploratory approach is an unusual procedure. Classically, we would postulate one or perhaps a handful of models, estimate some parameters, and then test for goodness of fit. Expediency and other practical considerations demand that we give up the classical procedures, and with them total objectivity and the right to make accurate probability statements about the statistical significance of our results. We have 125 cities to collect data on, and the Census provides thousands of measurements for each. Some of these measurements will appear related to others by chance alone, some are related by the way they were constructed, and others may not appear to be strongly collinear but still contain a hitherto unsuspected useful relationship. We need factual input of the type data alone can provide to direct future model building in useful directions. It is precisely this need that makes this study (and the Rand Urban Project as a whole) necessary.

Since many readers will be unfamiliar with the philosophy of exploratory data analysis, we first discuss a simple unrelated example. Suppose we wanted to forecast population totals for the United States in the next several years. Step one is to obtain some understanding of the history of the population totals, which might proceed as follows. Begin by plotting the data, which would expose an accelerating upward trend. That is not arguable—if a fixed proportion of the population reproduces, the magnitude of population increase will be proportional to the population. After converting population to logarithms, we might try a linear regression against time. Attention would be focused on the residuals from the regression. These turn out to be high in postwar years and low in recent times. Perhaps it would then occur to us that we should be looking at birth rates instead of population counts. We can smooth these birth rates (revealing empirical trends) or try to predict them from other considerations, and delve even deeper. Each expression of the data leads to ideas that permit a simpler or better expression of the data.

Turning to the cities, we found that the main problem is data reduction. Our eyes are overwhelmed by 125 cities and 53 variables (even carefully selected ones). Association on both coordinates of our 125 $\times$ 53 data matrix is a must. With respect to the variables, we have chosen factor analysis; with respect to the cities, cluster analysis. This is not because we especially believe the assumptions underlying the factor model or the cluster model, but because the output from these algorithms is plausible and simple. In addition, these methods are becoming standard to the field of urban demography, which makes comparisons with other work in the area easier. It should be noted that in our work, as in the other studies in this area, the relationships developed are between areas, not individuals, and small areas like Salinas are given equal weight with New York City. If different units—say, census blocks—were chosen, the results would be quite different.

FACTOR ANALYSIS

Factor analysis works on the supposition that many separate measurements can be well expressed in terms of a few unknown "factors." For mathematical readers, if V_{ij} is the ith measurment on city j,

$$V_{ij} = \sum_{p=1}^{f} \tilde{a}_{ip} \tilde{F}_{pj} + U_{ij} ,$$

where $\tilde{F}_{pj}$ = "factor score" for city j, factor p,
$\tilde{a}_{ip}$ = "factor loading" for variable i, factor p,
U_{ij} = residual, or unexplained part of variable i, city j,
f = number of factors.

There are many sophisticated criteria for a good fit for $\tilde{a}_{ip}$ and $\tilde{F}_{pj}$, but they essentially amount to this: We want to make the residuals $|U_{ij}|$ small without making the number of factors too large.

There is a useful indeterminacy in the formulation (1) because $\tilde{a}_{ip}$ and $\tilde{F}_{pj}$ are both created as part of the solution. To see what this is, rewrite (1) in matrix notation:

$$V = \widetilde{AF} + U \quad .$$

For any nonsingular $f \times f$ "rotation" (orthonormal) matrix R, we can re-express (2) as

$$V = (\tilde{A}R)(R'\tilde{F}) + U$$

$$= A\ F + U \quad .$$

In the factor analysis we used (program BMDO3M of the UCLA biomedical package), R was chosen to make the rows of F (factors) orthonormal and simultaneously to make the elements of A as close to zero or ± 1 as possible. The variables V_i (the rows of V) are initially adjusted to have unit length (giving each equal weight independent of the units of measurement). Also, the factors have unit length by construction, so A is restricted by

$$\sum_{p=1}^{f} a_{ip}^2 \leq 1 \quad .$$

In fact, the factor loading a_{ip} can be interpreted as the correlation between variable i and factor p.

Transformations of Variables

Behind the justification for factor analysis as it is currently practiced lies a hidden assumption of normality. If the data happen to contain extreme outliers, these outliers dominate the analysis. Special factors are created to explain only the outliers, and since the associated residuals are greatly reduced, the analysis appears to be functioning well. For variables where outliers seem to exist, we have selected a transformation to pull in those outliers. Almost always there is a transformation with some theoretical justification as well as desirable statistical properties, such as taking logarithms of population.

Variables having limited ranges in their natural form of expression also require transformation. Percentages are a frequent example of this. Without transformation, the difference between 1 percent and 2 percent is treated equally to the difference between 51 percent and 52 percent. A commonly used transformation for resolving this difficulty is:

$$S = \frac{1}{2} + \frac{1}{\Pi} \arcsin(2p-1)$$

where p is a percentage and S is the transformed value. If p is a binomial mean, then this is a "variance-stabilizing transformation." The action of this transformation is represented in the following table:

%	S	%	S
0	0.000	50	.500
1	.064	60	.565
2	.090	70	.631
5	.129	80	.705
10	.204	85	.747
15	.253	90	.796
20	.295	95	.871
30	.369	98	.910
40	.435	99	.936
50	.500	100	1.000

Our policy with regard to transformations can be summed up in one sentence: Since factor analysis is a linear method, one unit should be of equal importance in all ranges of the variable.

Selection of Variables: Redundancy

The selection of the input variables themselves has a crucial bearing on both the success and the interpretation of the factor analysis. When many different measures of a community facet (such as affluence) are available, there is a tendency to overrepresent them in the factor analysis. Why throw away useful information? But when we analyze the results, there is also a tendency to say that a factor that explains a lot of the variance is an important factor. This interpretation is a grave mistake, because the output variance is approximately proportional to the input variance. To see how this happens, consider a large factor analysis in which we add equivalent affluence measures one at a time. At some point an affluence factor will appear. After that, each new addition, and hence its variance, can be explained by the existing factor, which then increases in apparent importance. On the other hand, if the affluence factor explains the variance of other types of variables (such as climate or segregation), we have made a significant discovery. In this study we moderately limited the number of affluence variables input, and they were split up into other factors.

Conversely, a unique or an implicit variable may be buried because of underrepresentation. Unique variables can be statistically identified by having little of their variance explained by the important factors. A rough approximation to the variance explained by the key factors is the variance explained by the other variables, which is called the "estimated communality." It is printed as part of the factor analysis output. Implicit variables, however, have to be guessed. Consider the two variables, "central city income" and "suburban income." Both are measures of SMSA income, with suburban income generally higher. If we left these as is, we would learn nothing about city income vs. suburban income. First, we would have to examine many numbers even to get a quantitative idea of what the relationship looked like; and second, the fact that city-suburban differences are generally much smaller than inter-SMSA variation would be translated into a similar indication of importance by the factor analysis. This difficulty can be overcome by converting the city and suburban values into a difference and an average. Then the seemingly trivial practice of scaling all input variables to have unit variance has a crucial consequence: The small differences and large average values are scaled to the same size. We have included a number of implicit variables in this study.

For the purpose of dividing the cities into homogeneous groups, it is necessary to weight the factors by their estimated importance. Because of redundancy problems, and because the factors have been rotated for easy interpretation, there is no straightforward objective way of determining their weights. The weights have been based on the perceived importance of the class of variables that make up the factor, and on the variety of classes. (In some factors—such as age, welfare, or education—there is only one idea, whereas in others—such as inequality, poverty, and segregation—there are several.)

Some forms of data structure are not amenable to factor analysis. For example, quadratic and more general curvilinear relationships are not simple generalizations of linear ones. Multiple linear relationships, with cities of one type on one linear subspace and cities of another type in another subspace, could be completely unnoticed by a strictly linear method like factor analysis. We are continuing to explore various avenues of data representation at Rand.

CLUSTER ANALYSIS

Having assigned each city a score on each of the factors, we then distributed the cities into relatively homogeneous groups. This enables us to focus on individual cities without succumbing to tunnel vision. Moreover, it becomes easier to distinguish characteristics unique to one city from systematic differences. This is not to be confused with an assessment of importance, which we leave to urban experts.

The two most popular methods of cluster analysis, "top down" and "bottom up," are not well suited to the data. "Bottom up" begins with the 125 cities as 125 groups and coalesces groups based on the average (or perhaps minimum) distance between groups.[15] By the time one achieves ten clusters (say), there are two or three very large clusters and the rest contain one or two cities. On the other hand, "top down" starts with all the cities in a single cluster and successively divides them. The difficulty is that once we make a division we are stuck with it.

As an alternative, we have used a method that fixes the number of clusters and minimizes the sum of weighted squared distances of the points to their respective cluster centers. Starting from a random allocation, one looks at each city in sequence, assigning it to another cluster if that reduces the sum of weighted squared distances. The following well-known formulas make the checking easy:

Let

$$\bar{X}_n = \frac{1}{n} \Sigma X_i \ .$$

[15] The cities are represented by points in eight-dimensional space, with coordinates equal to their factor scores. The weighted distance between cities $x = (x_1, \ldots, x_8)$ and $y = (y_1, \ldots, y_8)$ is $(\Sigma(x_i - y_i)^2 \cdot W_i)^{1/2}$ where W_i is the subjective weight given to factor i.

Let

$$V_n = \frac{1}{n} \Sigma (X_i - \bar{X}_n)^2 .$$

Let

$$\Delta = X_{n+1} - \bar{X}_n .$$

Then

$$\bar{X}_{n+1} = \bar{X}_n + \Delta/(n+1)$$

$$V_{n+1} = V_n + (\frac{1}{n+1})\Delta^2 .$$

Convergence to the minimum does not always occur, although convergence to a relative minimum (one from which no single changes are profitable) is quite rapid. Although it would be theoretically pleasing to remove the dependence on a random initial allocation, it does not seem necessary since different starting points usually result in very similar clusterings.

Appendix C

OTHER CLASSIFICATIONS

People have been thinking about cities almost as long as they have been living in them. Earlier analysts have made numerous attempts at classification. In Table C-1, we show a few of these: by historical cycle of development, by function, by relation to other cities, or by economic base.[16]

Factor analytic classifications like ours do not use theory directly but attempt to describe as simply as possible the patterns that emerge from the data gathered. Thus, no classification can be absolutely general—the final pattern depends on what goes in. Other studies have concentrated on different types of variables, and their results provide a useful supplement to ours. In what follows, we will try to glean additional insights from the best recent factor analytic work.

Hadden and Borgatta collected 65 variables from the 1962 *City and County Data Book* and the 1960 Census. They split the towns over 25,000 into four groups by size. For cities over 150,000 they found ten major factors (those in parentheses were separate factors in the smaller-town analysis that were subsumed in the ten factors for large cities):

- Socioeconomic Status (Percent Nonwhite)
- Age Composition
- Educational Center
- Growth and Residential Mobility
- Density (Foreign Born, Public Transportation)
- Total Population
- Commercial Concentration (Wholesale, Retail, Manufacturing)
- Durables Manufacturing Concentration
- Unemployment
- Government Employees

The list is very similar to ours. Large cities that specialize in durables manufacturing have low education, less white-collar and other occupations; nondurables manufacturing co-exists with other types of commercial activities. The educational center

[16] Two excellent reviews of this field are J. K. Hadden and E. F. Borgatta, *American Cities: Their Social Characteristics,* Rand McNally, Chicago, 1965; and B. Berry (ed.), *City Classification Handbook.*

Table C-1

WAYS OF CLASSIFYING CITIES

	Historical Cycle	Functional	Economic Base[a]	Relation to Other Cities
Author	Mumford Forrester	Tower	Ogburn Harris Alexandersson	Kneedler
Main Categories	Primitive Developing Metropolis Megalopolis Chaos or stagnation	Commercial Industrial Political Recreational	Various manufacturing types Retailing Wholesaling Diversified Transport Mining Education Military Recreation Retirement, etc.	Independent Suburb Central city

[a]These studies, done mainly by geographers, classify all cities by city-forming activities--those economic activities (mining, furniture making, and the like) in which the city is more than 20 percent (A-type), 10-19 percent (B-type) or 5-9 percent (C-type) above the national average of employees for that activity.

factor is loaded mainly with "percent living in group quarters," which we did not collect. Socioeconomic status includes our education and inequality factors. In the United States, density is associated with foreign-born population (in the northeastern cities) and, because of the economies of concentration, with public transportation.[17] Criticizing the economic base studies, Hadden and Borgatta note that the numbers of persons employed in wholesale or retail trade correlate .9 or higher with population size and ask, "Does it make any sense to speak of cities specializing in wholesaling, retailing, manufacturing, etc. if the amount of each of these activities is directly proportional to the size of the city?" The economies of scale in the provision of public goods such as transportation, exotic restaurants, social services, and crime make size somewhat more important than it appears in our analysis where many of these effects are lessened by our use of rates.

Hadden and Borgatta do an interesting analysis of the stability of results under alternate definitions of urban area. The point is that there are three definitions of urban area used by the Census. The so-called Urbanized Area is exactly the densely settled part and may cut across various political boundaries. Within the urbanized

[17] For an interesting discussion of economies of concentration, as opposed to economies of scale, see M. Gaffney, "Containment Policies for Urban Sprawl," in *Approaches to the Study of Urbanization*, University of Kansas, Lawrence, 1970.

area is the legal central city, and the sum of all the counties that contain urbanized areas is the Standard Metropolitan Statistical Area (SMSA). Many of the western SMSAs contain great amounts of empty space, and in 1960 SMSA population was only 81 percent urban. Should we be worried that our results, based mainly on data collected for SMSAs, do not give a true picture of real urban patterns? Hadden and Borgatta made two tests. First, they correlated the entries in three national variable correlation matrixes, one for each definition of urban area. The SMSA variable correlations correlated .96 with those of the urban areas, .89 with those of the central cities, and the central cities correlated .92 with the urbanized areas. Thus all three definitions led to the same patterns with the urbanized areas falling between the two in the variables, just as it does physically. Second, they correlated the factor scores of the three definitions of the city—that is, they determined how similar the SMSAs, central cities, and urbanized areas of a certain area were, on the average. Of the factors, only density and population growth had correlations below .85 between any pair of definitions for the area. For most purposes, although it is important to make sure that the same definition is being used in each case of the cross-sectional analysis, it does not matter too much which definition it is.

An interesting contrast to the American work is given by Moser and Scott.[18] Reflecting England's greater homogeneity, they found that they could account for 60 percent of the variance of 60 variables with only four factors. These were social class, age of the area and growth 1931-1961, recent growth, and housing conditions. In Britain, status and demography are merged; the highest status communities are older, with smaller families and generally in exclusive suburbs or resorts. Their clustering presented three main groups: resorts and administrative and commercial centers, industrial towns, and suburbs. London was too distinct to be placed in any cluster.

Mayer's typology of 1960 SMSAs used 66 variables, mainly from the 1960 Census.[19] Five major factors emerged; socioeconomic status, age and size, stage in life cycle, recent growth, and nonmanufacturing. One interesting minor factor contained percent white, low rainfall, and elevation above sea level. Using these factors, he obtained the typology shown in Table C-2. Some differences between his work and ours are caused by the fact that he used all 212 SMSAs and more classes. Others are caused by our stress of suburban and city differences and neglect of size *per se*, and his subjective approach to classification. It is interesting to place his types into our clusters to see where the differences are. With the exception of our split of southern and big cities into those with differentiated and undifferentiated suburbs, all of Mayer's types are combined to form our clusters. For example, his Aa New England and C Mining towns combine to form our "declining white areas."

Meyer classified 145 SMSAs by characteristics of their nonwhite populations.[20] He found some interesting relationships between status and age reflected in the regions. First, the prosperous small northern industrial SMSAs had high status, young black families. In these cities, black males hold relatively good manufacturing

[18] C. A. Moser and W. Scott, *British Towns*, Oliver and Boyd, Edinburgh, 1961.

[19] H. M. Mayer, unpublished report cited by B. Berry and E. Neils, "Location, Size, and Shape of Cities," in H. S. Perloff (ed.), *The Quality of the Urban Environment*, Johns Hopkins Press, Baltimore, 1969.

[20] D. Meyer, "Classification of U.S. Metropolitan Areas by the Characteristics of Their Nonwhite Populations," in B. Berry (ed.), *City Classification Handbook*.

Table C-2

MAYER'S TYPOLOGY OF SMSAS, 1960

A. *New England, eastern New York, and New Jersey cities*
Intermediate to higher SES, older and/or larger, slow growth 1950-60, substantial commercial orientation, foreign born population, substantial use of public transport and cross-commuting.

Aa. *New England subgroups* (e.g.) Fall River, New Bedford)
Low status, older residual populations, crowding, etc.

Ab. *New York* (special case--modest status, old, large, commercial orientation, foreign born, public transport, etc.).

B. *Manufacturing belt cities*
Older and/or larger, industrial, slow growth 1950-60, high density, substantial foreign born, use of public transport.

C. *Mining towns* (Pennsylvania, West Virginia, Duluth)
Low SES, older populations, substantial use of group quarters, public transportation.

D. *Cities of agricultural Midwest and Plains*
Younger populations, slow growth 1950-60, commercial orientation, relative isolation, little use made of public transport.

Da. Chicago (special case--older, larger, manufacturing).

E. *Smaller towns of Pennsylvania, Ohio, Southern Indiana, and Border South*
Average or modest on all factors, few foreign born, somewhat older population, weaker commercial bases.

F. *Larger Mason-Dixon line cities, plus Atlanta, Richmond, Roanoke*
Some manufacturing, younger populations, slower growth, fewer foreign born.

G. *Southern cities*
Low SES, young populations, growing, weak commerce, few foreign born, substantial Negro population.

H. *Florida*
Older populations, rapid growth, commercial, many foreign born, relatively isolated, low density.

I. *Texas and Arizona*

Ia. *Texas Gulf coast*
Low density, substantial Negro populations and institutional or military base. Populations youngish, few foreign born.

Ib. *Mexican border towns*
Very low SES, very young populations, commercial, many foreign born, many institutional, military.

Ic. *West Texas and Arizona*
Higher SES, younger populations, very rapid growth, automobile-oriented, low density.

J. *Mountain States cities*
Young cities, young populations, commercial, few Negroes, relatively distant.

Ja. *Denver and Colorado Springs*
Same except larger, growing more rapidly, more use of public transport.

K. *West Coast cities*
Higher SES, commercial, substantial military involvement.

Ka. *Los Angeles* (special case--older, larger, more rapid growth, less commerce, absence of public transport).

L. *Other groups*

La. Principal "institutional" metropolitan areas--Ann Arbor, Champaign-Urbana, Lawton.

Lb. Las Vegas

Lc. Midland-Odessa

Ld. Honolulu

jobs, and their demographic characteristics are quite similar to their white blue-collar counterparts. The blacks in older declining industrial cities have high unemployment and older housing. The blacks in the largest cities are older but relatively prosperous. A high percentage of women head families and hold jobs. The west coast cities are distinguished by nonwhites with high education and high-status occupations. Many of these are Asian-Americans. The blacks in southern SMSAs are young, employed, but very poor. These are rural inmigrants. The job structure offers mainly low-paying jobs to blacks, but unlike in Texas, the blacks do not have to compete with anyone for them. The Texan and other southwestern cities have older but equally poor black populations with much higher unemployment. There is not as much rural inmigration, and although high-paying jobs are not available, there is competition with Mexicans and Indians for the low-prestige jobs.

Although there has been a plethora of city classifications, not much has been done with them. We next examine two studies of how useful classifications are in explanations.

Schnore and Winsborough tested the usefulness of Forstall's functional classification by degree of manufacturing against the manufacturing ratio itself in predicting suburban ghetto contrasts.[21] They give a simple regression of one measure of the contrasts.

$$\text{Income Suburb/City} = .24\ \text{Age City} - .29\ \text{City Population/Urban Area Population} + .17\ \text{City \% Black} + .19\ \text{Lacking Plumbing} + .31\ \text{Manufacturing Ratio}$$

In this case, dummies representing the classification add almost nothing to the regression. The standardized regression coefficients are interesting as each gives support to a different explanation of contrasts between suburbs and ghettos. Age of city is related to the developmental theory that new cities with the rich in the center age into those with some poor in the center and finally into those with only poor in the center. As central city housing ages, Tucson becomes Los Angeles and finally New York. The city population dominance may mean that it is too difficult to get away; the distance is too far, or the suburbs are not very well developed. The percent black in the central city can be interpreted as driving richer whites out, or, because of segregation, barring blacks from the suburbs. Lacking plumbing is a proxy for poor housing, which may be the cause or effect of fewer rich people in the city. Finally, the manufacturing ratio supports the theory that dirt, noise, and so on associated with manufacturing drive those who can afford it into the suburbs.

Clark tested factors against discrete variables in predicting some measures of political activity—League of Women Voters membership, reform government, decentralization, urban renewal, and general expenditures.[22] The final factor analysis predictions were never as good as those using discrete variables, because of the ambiguity of interpretation and muddling of effects. Nevertheless, he concludes that factors are useful. The primary reason is that they are orthogonal. Even when a

[21] L. F. Schnore and H. Winsborough, "Functional Classification and the Residential Location of Social Classes," in B. Berry (ed.), *City Classification Handbook;* R. L. Forstall, "Economic Classification of Places over 10,000, 1960," *The Municipal Yearbook 1967,* ICMA, Chicago, 1967.

[22] T. Clark, "Urban Typologies and Political Outputs," in B. Berry (ed.), *City Classification Handbook.*

great number of factors are included there are no multi-collinearity problems, and thus data exploration is more efficient. In addition, factors permit conclusions about classes of variables as opposed to this or that specific measure. Clark could claim that although socioeconomic characteristics are important in explanations of government outputs, the form of government and decentralization remain independently important.

Appendix D

CROSS SECTIONAL ANALYSIS: GROWTH AND DECLINE OF URBAN AREAS

To demonstrate how our data file can be used in testing theories we will present some preliminary work on growth and decline of urban areas. The city of St. Louis is losing population rapidly. A number of theories can be advanced to explain this phenomenon, but we were particularly concerned with the theory that the problem was an insufficient rate of economic growth for the area. We wanted to study the nationwide relationships between economic and population growth. Since the national employment rate is about 95 percent, it is difficult to separate the two in a given area; further, even if they are not moving to jobs as "recruited migrants," inmigrants generate retail, wholesale, and service jobs to cater to them. The correlation of area population growth to total income growth, as given by the *Survey of Current Business,* is .86. This mutual dependence and high collinearity led to a simultaneous equation model. The variables used in the model, with their means and standard deviations, are given in Table D-1.

The assumptions of the model are shown in Figure D-1, in which the arrows indicate influence. For example, it is assumed that SMSA total income growth and population growth are jointly determined, and that SMSA population growth together with other exogenous variables determine central city population growth.

The results for 124 urban areas[23] and for the 59 largest cities are shown in Table D-2.[24] For the 124 areas case, income growth and SMSA population growth are closely linked, but the South and other poor and poorly educated areas are catching up in income. This leveling is to be expected as national influences become more important on local areas. In addition, we see that Congressional power, stronger city governments, and manufacturing have added to income growth. Natural increase and a good climate have an independent effect on population growth. There are two reasons why poorer areas were catching up in income in the decade. Poor people continued to migrate out, and many of those who stayed improved their relative position. Unfortunately, this model does not allow us to assess the relative importance of the reasons. Central city change is mainly influenced by SMSA population

[23] Honolulu appears to be a special case and was dropped from the analysis.

[24] Because the model is a system of simultaneous equations, two-stage least squares was used to estimate the parameters.

Table D-1

SELECTED STATISTICS FOR 124 METROPOLITAN AREAS[a]

Variable Number	Name	Abbreviation	Mean	Standard Deviation
	Total Income Growth[b]	EcGro	.0709	.015
4	SMSA Population Change[b]	SMSCh	.0175	.014
	Central City Population Change[b]	CC Ch	.0070	.019
	Congressional Power[c]	CONG P	.24	.45
49	Type of City Government	C GOV	2.60	1.30
14	Manufacturing Ratio[d]	MANUF	.34	.08
17	High School Graduates[d]	HSG	.53	.05
	Age of City[e]	Age C	2.03	.31
8	Gini Coefficient	Gini	.341	.028
52	Federal Employees[d]	FEDEM	.07	.03
	South (Dummy)	SOUTH	.32	.47
	Birth Rate minus Death Rate, 1968	Nat Inc	.0089	.003
23	Unemployment, 1970[d]	Unemp	.13	.018
9	Climate	Clim	-6.8	9.5
	Old in Central City, 1960 (decile)	CC Old	4.9	2.5
	Density Central City (log)	DENSC	3.73	.26
	Black Central City, 1960[d]	Black	.24	.12

[a]Data Base minus Honolulu, which doesn't seem to fit into the same pattern.

[b]Annual rate of growth, 1960-1970.

[c]This variable is 1 if a key congressman or senator has city as a base.

[d]Transformed by variance preserving transformation (see Appendix B).

[e]The population of the city in (1890+1910+1930)/1960.

change, but older cities with more old or black citizens lost more population, even with SMSA population change taken into account.

The story is different when only cities over 200,000 population are considered. The great size of these areas is important to interpretation. They are much more alike than smaller areas. Since they are generally the complete world of their citizens, they provide in a determined way the necessary range of economic services. Their size makes change more difficult. Thus, Congressional power is apparently diluted to insignificance, and city government, manufacturing ratio, and federal employees also lose significance. Climate loses its importance in predicting SMSA population change, and age of city is less important in predicting central city change. SMSA population growth, the one force big enough to make a difference, appears to have a multiplier effect on income growth. With these exceptions, the basic pattern remains the same.

What do these results imply about central city decline in St. Louis? Using the 124 city regression results, we can estimate the effect of different St. Louis characteristics on the growth rate. The city has been losing population at a 2 percent annual

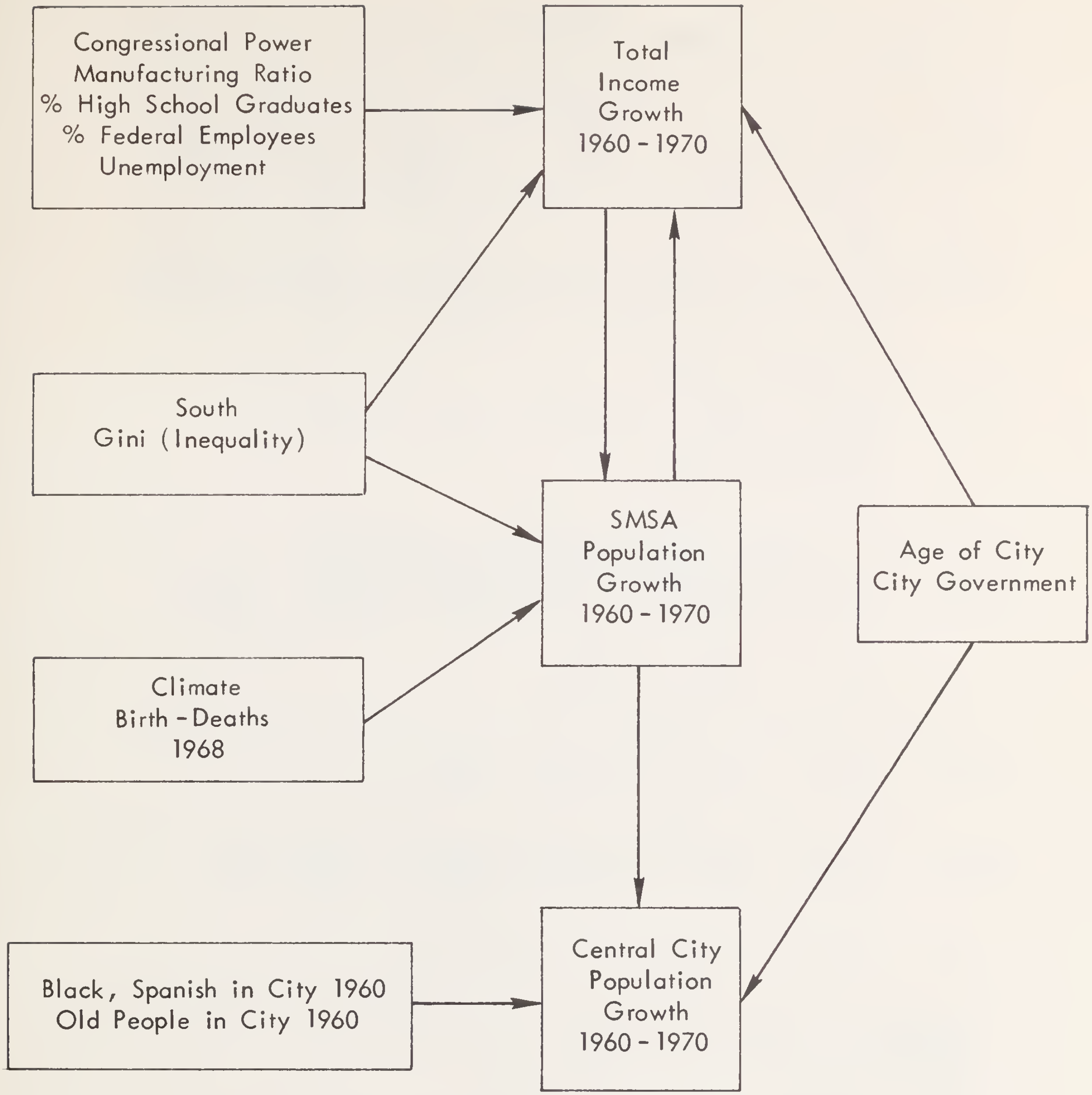

Fig. D-1—Model of metropolitan growth

Table D-2

A MODEL OF GROWTH AND DECLINE[a]

124 Urban Areas[b]

EcGro[c] =	.95 SMSCh	+ .0035 Cong P	+ .0015 C Gov	+ .025 Manuf	− .023 HSG	+ .0072 Age C
	(10)[d]	(2.9)	(3.5)	(1.9)	(1.4)	(2.2)
	+ .082 Gini	+ .0055 South	+ .017 Fedem	− .039 Unemp	+ .13	
	(2.3)	(2.8)	(.7)	(1.1)	(.6)	

R^2 = .85 S.E. = .00565

SMSCh =	.93 EcGro	− .042 Gini	− .007 South	+ .37 Nat Inc	+ .0002 Clim	− .0342
	(11)	(+1.2)	(3.4)	(1.9)	(2.25)	(2.2)

R^2 = .80 S.E. = .0063

CC Ch =	.60 SMSCh	− .018 Age C	− .026 Black	+ .001 C Gov	− .0012 CC Old
	(3.9)	(2.7)	(+2.3)	(1.1)	(2.3)
	+ .0084 Densc	+ .0114			
	(1.7)	(.5)			

R^2 = .63 S.E. = .0114

59 Urban Areas with Central Cities Over 200,000[d]

EcGro =	1.59 SMSCh	+ .018 Manuf	− .073 HSG	+ .01 Age C	+ .19 Gini	− .0048 South
	(6.1)	(.7)	(2.1)	(1.9)	(2.4)	(1.3)
	+ .34 Fedem	− .082 Unemp	+ .0003			
	(.8)	(1.2)	(.007)			

R^2 = .75 S.E. = .0065

SMSCh =	.82 EcGro	− .0038 South	+ .28 Nat Inc	− .042
	(7.4)	(2.2)	(1.2)	(5.2)

R^2 = .75 S.E. = .0052

CC Ch =	.73 SMSCh	− .007 Age C	− .026 Black	+ .0018 C Gov	− .0017 CC Old
	(2.9)	(.81)	(1.6)	(1.4)	(2.2)
	+ .0057 Densc	− .0043			
	(.9)	(.15)			

R^2 = .59 S.E. = .0103

[a]Estimated by two-stage least squares.

[b]All but Honolulu.

[c]See Table D-1 for explanation of abbreviations.

[d]Values in parentheses are t-ratios. Four variables with t-ratios less than .5 are not listed. These variables are Congressional Power, Type of City Government in the top equation, and Climate and the Gini Coefficient in the second equation.

rate, 3 points lower than the 124 city average annual gain of 1 percent. The model predicts slower growth for the area as a whole because of the lack of Congressional power, the weak city government, high rates of unemployment, and rather poor climate. In fact, the SMSA growth rate is .55 percent less than the national average. The effect of this on the annual rate of city growth is estimated to be $.6 \times -.55\% = -.33\%$. The other major influences according to the model are St. Louis' age (−.7%), high % black in 1960 (−.3%), weak city government (−.2%), high median age in 1960 (−.2%), and high density (−.2%). Adding all these effects, we see that St. Louis is predicted to have an annual growth rate of about −1%, a slight overestimate. Low economic growth accounts for only a small part of the central city decline. Indeed, the analysis reinforces the point that most of the major determinants of big city problems are not controllable by local officials, or, in fact, by anyone.

BIBLIOGRAPHY

Alexandersson, G., *The Industrial Structure of American Cities,* University of Nebraska Press, Lincoln, 1956.

Alford, R., "Critical Evaluation of the Principles of City Classification," in B. Berry (ed.), 1972.

Berry, B. (ed.), *City Classification Handbook: Methods and Applications,* Wiley, New York, 1972.

—— and E. Neils, "Location, Size and Shape of Cities," in H. S. Perloff (ed.), 1969.

Clark, T., "Urban Typologies and Political Outputs," in B. Berry (ed.), 1972.

Forrester, J. W., *Urban Dynamics,* MIT Press, Cambridge, 1969.

Forstall, R. L., "Economic Classification of Places over 10,000, 1960," *The Municipal Year Book 1967,* International City Managers' Association, Chicago, 1967.

Gaffney, M., "Containment Policies for Urban Sprawl," in *Approaches to the Study of Urbanization,* University of Kansas, Lawrence, 1970.

Goldberger, A., *Econometric Theory,* Wiley, New York, 1964.

Grossman, M., "The Demand for Health," Occasional Paper 119, National Bureau of Economic Research, Columbia Press, New York, 1972.

Hadden, J. K. and E. F. Borgatta, *American Cities: Their Social Characteristics,* Rand McNally, Chicago, 1965.

Harris, C. D., "A Functional Classification of Cities in the United States," *Geographical Review,* 33, January 1943, 86-99.

Kneedler, G., "Economic Classification of Cities," *Municipal Yearbook, 1945,* International City Managers' Association, Chicago, 1945.

Mayer, H. M., unpublished report cited in Berry and Neils, in H. S. Perloff (ed.), 1969.

Meyer, D., "Classification of U.S. Metropolitan Areas by the Characteristics of Their Nonwhite Populations," in B. Berry (ed.), 1972.

Moser, C. A. and W. Scott, *British Towns,* Oliver and Boyd, Edinburgh, 1961.

Mumford, L., *The Culture of Cities,* Harcourt, New York, 1938.

Ogburn, W. F., *Social Characteristics of Cities,* International City Managers' Association, Chicago, 1937.

Perloff, H. S. (ed.), *The Quality of the Urban Environment,* Johns Hopkins Press, Baltimore, 1969.

Pickard, P., *U.S. Metropolitan Growth and Expansion 1970-2000 with Population Projections,* Urban Land Institute, Washington, 1971.

Schnore, L. F. and H. Winsborough, "Functional Classification and the Residential Location of Social Classes," in B. Berry (ed.), 1972.

Thorndike, E. L., *Your City,* Harcourt, New York, 1939.

Tower, W. D., "The Geography of American Cities," *Bulletin of the American Geographic Society,* 37, 1905, 577-588.

U.S. Bureau of the Census, *County and City Data Book,* 1967.

——, *U.S. Census of Population: 1970 General Social and Economic Characteristics.*